Manuale dell'utente del modulo SPSS Statistics Base 17.0

Per ulteriori informazioni sui prodotti software SPSS Inc., visitare il sito Web all'indirizzo *http://www.spss.it* o contattare:

SPSS Inc.
233 South Wacker Drive, 11th Floor
Chicago, IL 60606-6412
Tel: (312) 651-3000
Fax: (312) 651-3668

SPSS è un marchio registrato e gli altri nomi di prodotti sono marchi di software di proprietà di SPSS Inc.. Non è consentito produrre o distribuire materiale che descriva il software senza previa autorizzazione scritta dei proprietari dei diritti di marchio e dei diritti di licenza per il software, nonché dei diritti di copyright per la documentazione.

Il SOFTWARE e la documentazione vengono forniti con DIRITTI LIMITATI. L'utilizzo, la duplicazione e la divulgazione da parte del Governo sono soggetti alle restrizioni indicate nella sottoclausola (c)(1)(ii) della clausola The Rights in Technical Data and Computer Software al punto 52.227-7013. Il contraente/produttore è SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.
Brevetto n. 7.023.453

Nota: altri nomi di prodotti menzionati nel presente documento vengono utilizzati esclusivamente a fini identificativi e sono marchi dei rispettivi proprietari.

Windows è un marchio registrato di Microsoft Corporation.

Apple, Mac e il logo Mac sono marchi registrati di Apple Computer, Inc., registrato negli Stati Uniti e in altri paesi.

Questo prodotto utilizza WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, *http://www.winwrap.com*.

Nessuna parte di questa pubblicazione potrà comunque essere riprodotta o inserita in un sistema di riproduzione o trasmessa in qualsiasi forma e con qualsiasi mezzo (in formato elettronico, meccanico, su fotocopia, come registrazione o altro) per qualsiasi scopo, senza il permesso scritto dell'editore.

Prefazione

SPSS Statistics 17.0

SPSS Statistics 17.0 è un sistema completo per l'analisi dei dati. SPSS Statistics è in grado di elaborare dati provenienti da quasi tutti i tipi di file e di utilizzarli per generare rapporti con tabelle, grafici, grafici di distribuzioni e trend, statistiche descrittive e complesse analisi statistiche.

Questo manuale, *Manuale dell'utente del modulo SPSS Statistics Base 17.0*, fornisce indicazioni sull'interfaccia grafica di SPSS Statistics. Nell'Aiuto in linea installato con il software sono forniti esempi relativi all'utilizzo delle procedure statistiche disponibili in SPSS Statistics Base 17.0.

Oltre ai menu e alle finestre di dialogo, in SPSS Statistics è possibile utilizzare un linguaggio a comandi. Alcune caratteristiche avanzate del sistema possono essere utilizzate soltanto mediante la sintassi dei comandi. Tali caratteristiche non sono disponibili nella versione per studenti. Informazioni dettagliate sui riferimenti alla sintassi dei comandi sono disponibili in due formati: nella Guida integrate e su un documento PDF a parte in *Command Syntax Reference*, anch'esso consultabile tramite il menu ?.

SPSS Statistics Opzioni

Le seguenti opzioni sono disponibili come componenti avanzati aggiuntivi per il sistema completo SPSS Statistics Base (non per la versione per studenti):

Regression fornisce tecniche per l'analisi dei dati che non rientrano nei modelli statistici lineari tradizionali. Include procedure per l'analisi Probit, la regressione logistica, il metodo dei minimi quadrati ponderati, la regressione con il metodo dei minimi quadrati ponderati e la regressione non lineare generalizzata.

Advanced Statistics offre tecniche spesso utilizzate nella ricerca biomedica e sperimentale avanzata. Include procedure per modelli lineari generalizzati (GLM), modelli misti lineari, analisi della varianza, analisi loglineare, regressione ordinale, tavole di sopravvivenza attuariali, analisi della sopravvivenza di Kaplan-Meier e regressione di Cox di base ed estesa.

Custom Tables consente di creare una vasta gamma di rapporti con tabelle, incluse le tabelle complesse con intestazioni e colonne, e consente di visualizzare dati a risposta multipla.

Forecasting consente di eseguire previsioni complete e analisi di serie storiche con più modelli per la stima delle curve, modelli di livellamento e metodi per la stima delle funzioni di autoregressione.

Categories consente di eseguire procedure di scaling ottimale, inclusa l'analisi delle corrispondenze.

Conjoint consente di misurare in modo realistico l'influenza degli attributi di prodotti individuali sulle preferenze dei consumatori e dei cittadini. Mediante Conjoint, è possibile misurare in modo semplice l'effetto dello scambio di ogni attributo di prodotti nel contesto di un insieme di attributi di prodotti —e verificare la reazione dei consumatori all'atto di compiere una decisione di acquisto.

Exact Tests consente di calcolare valori *p* esatti nei test statistici, soprattutto quando, a causa di campioni piccoli o mal distribuiti, i test comuni potrebbero risultare non corretti. Questa opzione è disponibile solo nei sistemi operativi Windows.

Missing Values descrive i modelli dei dati mancanti, fornisce una stima delle medie e di altre statistiche e assegna i valori per le osservazioni mancanti.

Complex Samples consente ai ricercatori che eseguono sondaggi, indagini di mercato, ricerche in ambito medico o dell'opinione pubblica, nonché agli scienziati che si avvalgono della metodologia basata sui sondaggi campione, di incorporare nell'analisi dei dati i disegni di campioni complessi creati.

Decision Trees crea un modello di classificazione basato su alberi. Classifica i casi in gruppi o prevede i valori di una variabile dipendente (di destinazione) in base ai valori di variabili (predittore) indipendenti. La procedura offre strumenti di validazione per l'analisi di classificazione confermativa ed esplorativa.

Data Preparation fornisce un'istantanea immediata dei dati. In particolare consente di applicare le regola di convalida che identificano i valori di dati non validi. È possibile creare regole che escludono i valori fuori intervallo, i valori mancanti o vuoti. È possibile anche salvare le variabili che registrano violazioni alle singole regole nonché il numero totale di violazioni alle regole per caso. Viene fornito un set limitato di regole predefinite da copiare o modificare.

Neural Networks può essere utilizzato per prendere decisioni aziendali tramite la previsione della domanda relativa a un prodotto in funzione del prezzo e di altre variabili oppure classificando i clienti in base alle abitudini di acquisto e alle caratteristiche demografiche. Le reti neurali sono strumenti per la creazione di modelli basati su dati non lineari. Possono essere utilizzate per creare modelli di relazioni complesse tra gli input e gli output o per individuare motvi nei dati.

EZ RFM esegue l'analisi RFM (data ultimo acquisto, frequenza, valore monetario) sui file di dati delle transazioni e dei clienti.

Amos™ (**a**nalysis of **mo**ment **s**tructures) utilizza modelli di equazioni strutturali per confermare e spiegare modelli concettuali che riguardano gli atteggiamenti, le percezioni e altri fattori comportamentali.

Installazione

Per installare Base modulo, eseguire l'Attivazione guidata licenza utilizzando il codice di autorizzazione fornito da SPSS Inc.. Per ulteriori informazioni, consultare le istruzioni di installazione fornite con Base modulo.

Compatibilità

SPSS Statistics è progettato per l'esecuzione in una vasta gamma di sistemi operativi. Per informazioni specifiche sui requisiti minimi e consigliati, vedere le istruzioni di installazione fornite con il sistema.

Numeri di serie

Il numero di serie è il numero di identificazione del cliente per SPSS Inc.. Sarà necessario fornire questo numero nel caso l'utente contatti SPSS Inc. per ricevere informazioni relative al supporto, al pagamento e ai sistemi aggiornati. Il numero di serie viene fornito con il modulo Base.

Servizio clienti

Per informazioni sulla spedizione o sul proprio account, contattare la filiale SPSS nel proprio paese indicata nell'elenco disponibile nel sito Web all'indirizzo *http://www.spss.com/worldwide/*. Tenere presente che sarà necessario fornire il numero di serie.

Corsi di formazione

SPSS Inc. organizza corsi di formazione pubblici e onsite che includono esercitazioni pratiche. Tali corsi si terranno periodicamente nelle principali città. Per ulteriori informazioni sui corsi, contattare la filiale SPSS nel proprio paese indicata nell'elenco disponibile nel sito Web all'indirizzo *http://www.spss.com/worldwide/*.

Supporto tecnico

Ai clienti che richiedono la manutenzione, viene messo a disposizione un servizio di supporto tecnico. I clienti possono contattare il supporto tecnico per richiedere assistenza sull'utilizzo dei prodotti SPSS Statistics o sull'installazione di uno degli ambienti hardware supportati. Per il supporto tecnico, visitare il sito Web di SPSS all'indirizzo *http://www.spss.it* o contattare la filiale SPSS nel proprio paese indicata nel sito Web all'indirizzo *http://www.spss.com/worldwide*. Tenere presente che sarà necessario fornire dati di identificazione personali e relativi alla propria società e il numero di serie del sistema.

Pubblicazioni aggiuntive

Presso Prentice Hall è disponibile il volume *SPSS Statistical Procedures Companion* di Marija Norusis, che sarà presto disponibile in una versione aggiornata per SPSS Statistics 17.0. Verrà preseto pubblicato anche il volume *SPSS Advanced Statistical Procedures Companion*, basato su SPSS Statistics 17.0. A breve verrà sviluppata anche la pubblicazione *SPSS Guide to Data Analysis* for SPSS Statistics 17.0. I titoli dei volumi disponibili esclusivamente tramite Prentice Hall verranno pubblicati sul sito Web all'indirizzo *http://www.spss.com/estore* (selezionare il proprio paese di residenza, quindi fare clic su Books).

Contenuto

1 Osservazioni generali 1

Novità della versione 17.0 . . . 2
Finestre . . . 4
Confronto tra finestra designata e attiva . . . 5
Barra di stato . . . 6
Finestre di dialogo . . . 6
Nomi ed etichette di variabili negli elenchi delle finestre di dialogo . . . 6
Ridimensionamento delle finestre di dialogo . . . 7
Controlli delle finestre di dialogo . . . 8
Selezione di variabili . . . 8
Icone relative al tipo di dati, al livello di misurazione e all'elenco delle variabili . . . 8
Per ottenere informazioni sulle variabili in una finestra di dialogo . . . 9
Passaggi fondamentali dell'analisi dei dati . . . 9
Statistics Coach . . . 10
Ulteriori informazioni . . . 10

2 Aiuto 11

Visualizzazione di informazioni sui termini delle tabelle pivot dell'output . . . 12

3 File di dati 14

Apertura di file di dati . . . 14
Per aprire file di dati . . . 14
Tipi di file di dati . . . 15
Apertura file: opzioni . . . 15
Lettura di file di Excel 95 o versione successiva . . . 16
Lettura di file Excel di versioni precedenti e di altri fogli elettronici . . . 16
Lettura di file dBASE . . . 17
Lettura di file Stata . . . 17
Lettura di file di database . . . 17
Importazione guidata di testo . . . 32
Lettura di dati Dimensions . . . 41
Informazioni sul file . . . 44

Salvataggio di file di dati 45
Salvataggio di file di dati modificati 45
Salvataggio di file di dati in formati esterni 45
Salvataggio di file di dati in formato Excel 48
Salvataggio di file di dati in formato SAS 49
Salvataggio di file di dati in formato Stata 50
Salvataggio di sottoinsiemi di variabili 51
Esportazione a un database 52
Esportazione a Dimensions 64
Protezione dei dati originali 65
File attivo virtuale 65
Creazione di una copia dei dati 67

4 Analisi distribuita 69

Login al server 69
Aggiunta e modifica delle impostazioni di login al server 70
Per selezionare, cambiare o aggiungere server 71
Ricerca dei server disponibili 72
Apertura di file di dati da un server remoto 73
Accesso ai file in modalità di analisi locale e distribuita 73
Procedure disponibili in modalità di analisi distribuita 75
Specifiche di percorsi assoluti e relativi 75

5 Editor dei dati 77

Visualizzazione dati 77
Visualizzazione variabili 78
Per visualizzare e definire gli attributi delle variabili 79
Nomi delle variabili 79
Livello di misurazione delle variabili 80
Tipo di variabile 81
Etichette delle variabili 83
Etichette dei valori 84
Inserimento di interruzioni di riga nelle etichette 84
Valori mancanti 85
Larghezza colonna 86
Allineamento 86
Applicazione degli attributi di definizione di una variabile ad altre variabili 86
Attributi variabili personalizzati 87

Personalizzazione della Visualizzazione variabili . . . 91
Controllo ortografico . . . 92
Inserimento di dati . . . 92
Per inserire dati numerici . . . 93
Per inserire dati non numerici . . . 93
Per utilizzare etichette dei valori per l'inserimento dei dati . . . 93
Limitazioni relative ai valori . . . 94
Modifica dei dati nella Visualizzazione dati . . . 94
Sostituzione o modifica di valori . . . 94
Tagliare, copiare e incollare valori . . . 94
Inserimento di nuovi casi . . . 95
Inserimento di nuove variabili . . . 96
Per modificare il tipo di dati . . . 96
Ricerca di casi, variabili o assegnazioni . . . 97
Ricerca e sostituzione dei valori dei dati e degli attributi . . . 98
Criteri di selezione dei casi nell'Editor dei dati . . . 99
Editor dei dati: Opzioni di visualizzazione . . . 100
Stampa dall'Editor dei dati . . . 101
Per stampare il contenuto dell'Editor dei dati . . . 101

6 Utilizzo di più sorgenti dati 102

Gestione di più sorgenti di dati . . . 103
Utilizzo di più insiemi di dati nella sintassi dei comandi . . . 104
Copia e inserimento di informazioni in più insiemi di dati . . . 105
Ridenominazione degli insiemi di dati . . . 105
Soppressione di più insiemi di dati . . . 106

7 Preparazione dei dati 107

Proprietà variabili . . . 107
Definizione di proprietà delle variabili . . . 108
Per definire le proprietà delle variabili . . . 108
Definizione di etichette di valori e di altre proprietà delle variabili . . . 109
Assegnazione del livello di misurazione . . . 111
Attributi variabili personalizzati . . . 112
Copia di proprietà delle variabili . . . 113
Insiemi a risposta multipla . . . 114
Definizione di insiemi a risposta multipla . . . 114

Copia delle proprietà dei dati 117
Per copiare le proprietà dei dati 118
Selezione delle variabili sorgente e destinazione 118
Selezione delle proprietà delle variabili da copiare 120
Copia delle proprietà dell'insieme di dati (file) 121
Risultati 124
Identificazione di casi duplicati 124
Categorizzazione visuale 127
Per categorizzare le variabili 128
Categorizzazione di variabili 128
Generazione automatica di categorie raggruppate 131
Copia di categorie raggruppate 133
Valori mancanti definiti dall'utente in Categorizzazione visuale 134

8 Trasformazioni di dati 135

Calcolo delle variabili 135
Calcola variabile: Condizioni 137
Calcola variabile: Tipo ed etichetta 137
Funzioni 138
Valori mancanti nelle funzioni 138
Generatori di numeri casuali 139
Conteggia occorrenze di valori all'interno dei casi 140
Conteggia valori all'interno dei casi: Valori da conteggiare 140
Conta ricorrenze: Condizioni 141
Sposta valori 142
Ricodifica dei valori 143
Ricodifica nelle stesse variabili 143
Ricodifica nelle stesse variabili: Valori vecchi e nuovi 144
Ricodifica in variabili differenti 146
Ricodifica in variabili differenti: Valori vecchi e nuovi 147
Ricodifica automatica 148
Classifica casi 151
Classifica casi: Tipi 152
Classifica casi: Pari merito 153
Impostazione guidata data e ora 154
Date e ore in SPSS Statistics 156
Creazione di una variabile data/ora da una stringa 156
Creazione di una variabile data/ora da un insieme di variabili 158
Aggiunta o sottrazione di valori dalle variabili data/ora 160

Estrazione di parte della variabile data/ora 167
Trasformazioni di dati di serie storiche 169
Definisci date 170
Crea serie storica 171
Sostituisci valori mancanti 174
Calcolo del punteggio dei dati con modelli predittivi 175
Caricamento di un modello salvato 176
Visualizzazione di un elenco di modelli caricati 179
Funzionalità aggiuntive disponibili con la sintassi dei comandi 179

9 Gestione e trasformazione di file 180

Ordina casi 180
Ordina variabili 181
Trasponi 182
Unione di file di dati 183
Aggiungi casi 183
Aggiungi casi: Rinomina 186
Aggiungi casi: informazioni del dizionario 186
Unione di più sorgenti dati 186
Aggiungi variabili 186
Aggiungi variabili: Rinomina 188
Unione di più sorgenti dati 188
Aggrega dati 188
Aggrega dati: Funzione di aggregazione 191
Aggrega dati: Etichetta e nome variabile 192
Dividi 192
Seleziona casi 193
Seleziona casi: Se 195
Seleziona casi: Campione casuale 196
Seleziona casi: Intervallo 196
Pesa casi 197
Ristrutturazione di dati 198
Per ristrutturare i dati 198
Ristrutturazione di dati guidata: Seleziona tipo 199
Ristrutturazione di dati guidata (Variabili in casi): Numero di gruppi di variabili 202
Ristrutturazione di dati guidata (Variabili in casi): Seleziona Variabili 203
Ristrutturazione di dati guidata (Variabili in casi): Crea variabili Variabili 205
Ristrutturazione di dati guidata (Variabili in casi): Crea una variabile indice 207
Ristrutturazione di dati guidata (Variabili in casi): Crea più variabili indice 208
Ristrutturazione di dati guidata (Variabili in casi): opzioni 209

Ristrutturazione di dati guidata (Casi in variabili): Seleziona Variabili 210
Ristrutturazione di dati guidata (Casi in variabili): Ordinamento Dati 212
Ristrutturazione di dati guidata (Casi in variabili): opzioni 212
Ristrutturazione di dati guidata: Fine 214

10 Utilizzo dell'output 216

Viewer 216
Attivazione e disattivazione della visualizzazione dell'output 217
Spostamento, copia ed eliminazione dell'output 217
Modifica dell'allineamento iniziale 218
Modifica dell'allineamento degli elementi dell'output 218
Vista riassuntiva del Viewer 218
Aggiunta di elementi nel Viewer 220
Individuazione e sostituzione di informazioni nel Viewer 221
Copia dell'output in altre applicazioni 222
Per copiare e incollare gli elementi di output in un'altra applicazione 223
Esporta output 223
Opzioni HTML 225
Opzioni Word/RTF 226
Opzioni Excel 227
Opzioni di PowerPoint 229
Opzioni PDF 231
Opzioni testo 232
Opzioni solo grafica 233
Opzioni di formato grafico 234
Stampa di documenti del Viewer 235
Per stampare output e grafici 235
Anteprima di stampa 236
Attributi di pagina: intestazioni e piè di pagina 237
Attributi di pagina: Opzioni 238
Salvataggio dell'output 239
Per salvare un documento Viewer 239

11 Tabelle pivot 241

Gestione di una tabella pivot 241
Attivazione di una tabella pivot 241
Pivoting di una tabella 241
Modifica dell'ordine di visualizzazione degli elementi all'interno di una dimensione 242

Spostamento di righe e colonne all'interno di un elemento relativo alle dimensioni 242
Per trasporre righe e colonne 243
Raggruppamento di righe o colonne 243
Separazione di righe o colonne 243
Rotazione delle etichette di riga o colonna 243
Utilizzo degli strati 244
Creazione e visualizzazione di strati 244
Vai a categoria strato 246
Visualizzare o nascondere gli elementi 247
Nascondere le righe e le colonne di una tabella 247
Visualizzazione di righe e colonne nascoste in una tabella 247
Nascondere e visualizzare le etichette delle dimensioni 247
Nascondere e visualizzare i titoli delle tabelle 248
Modelli di tabelle 248
Per applicare o salvare un TableLook 248
Per modificare o creare un TableLook 249
Proprietà tabella 249
Per modificare le proprietà delle tabelle pivot 250
Proprietà tabella: Generale 250
Proprietà tabella: Didascalie 251
Proprietà tabella: Formati di cella 252
Proprietà tabella: Bordi 254
Proprietà tabella: Stampa 255
Proprietà cella 256
Carattere e sfondo 257
Formatta valore 257
Allineamento e margini 258
Annotazioni e didascalie 259
Aggiunta di annotazioni e didascalie 259
Per visualizzare o nascondere una didascalia 260
Per nascondere o visualizzare un'annotazione in una tabella 260
Rimando a nota 260
Per rinumerare le annotazioni 261
Larghezza delle celle 261
Modifica della larghezza di colonna 261
Visualizzazione dei bordi nascosti in una tabella pivot 261
Selezione di righe e colonne in una tabella pivot 262
Stampa delle tabelle pivot 263
Impostazione delle interruzioni di pagina per le tabelle con larghezza o lunghezza eccessiva 263
Creazione di un grafico da una tabella pivot 264

12 Modelli 265

Interazione con un modello . . . 265
Uso del Viewer modelli . . . 265
Stampa di un modello . . . 267
Esportazione di un modello . . . 267

13 Utilizzo della sintassi dei comandi 268

Regole della sintassi dei comandi . . . 268
Creazione di sintassi dei comandi nelle finestre di dialogo . . . 270
Per creare sintassi dei comandi nelle finestre di dialogo . . . 270
Copia della sintassi dal registro dell'output . . . 271
Per copiare la sintassi dal registro dell'output . . . 271
Utilizzo dell'Editor della sintassi . . . 272
Finestra Editor della sintassi . . . 273
Terminologia . . . 274
Autocompletamento . . . 275
Codifica colori . . . 275
Punti di interruzione . . . 276
Segnalibri . . . 278
Impostazione del testo come commento . . . 279
Esecuzione della sintassi dei comandi . . . 279
File di sintassi Unicode . . . 281
Comandi EXECUTE multipli . . . 281

14 Informazioni sui dati 282

Scheda Output della finestra Informazioni sui dati . . . 284
Scheda Informazioni sui dati - Statistiche . . . 287

15 Frequenze 289

Frequenze: Statistiche . . . 290
Frequenze: Grafici . . . 292
Frequenze: Formato . . . 293

16 Descrittive 294

Descrittive: Opzioni. 295
Funzioni aggiuntive del comando DESCRIPTIVES 296

17 Esplora 298

Esplora: Statistica 299
Esplora: Grafici. 300
Esplora: potenza necessaria per la trasformazione dei dati 301
Esplora: Opzioni 302
Funzioni aggiuntive del comando EXAMINE 302

18 Tavole di contingenza 303

Strati nelle tavole di contingenza 304
Tavole di contingenza: Grafici a barre raggruppati 305
Tavole di contingenza: Statistiche. 305
Tavole di contingenza: Visualizzazione cella 308
Tavole di contingenza: Formato tabella 309

19 Riassumi 310

Riassumi: Opzioni 312
Riassumi: Statistiche 312

20 Medie 315

Medie: Opzioni 317

21 Cubi OLAP 320

Cubi OLAP: Statistiche 321

Cubi OLAP: Differenze 324
Cubi OLAP: Titolo 325

22 Test T 326

T per campioni indipendenti 326
Test T per campioni indipendenti: Definisci gruppi 328
Test T per campioni indipendenti: Opzioni 328
T per campioni appaiati 329
Test T per campioni appaiati: Opzioni 330
Test T per un campione 331
Test T per un campione: Opzioni 332
Opzioni aggiuntive del comando T-TEST 332

23 ANOVA univariata 333

ANOVA univariata: Contrasti 334
ANOVA univariata: Test Post Hoc 335
ANOVA univariata: Opzioni 337
Opzioni aggiuntive del comando ONEWAY 338

24 Analisi GLM univariato 339

GLM - Univariato: Modello 341
Costruisci termini 341
Somma dei quadrati 342
GLM - Univariato: Contrasti 343
Tipi di contrasto 343
GLM - Univariato: Profili 344
GLM - Univariato: Confronti post hoc 345
GLM - Univariato: Salva 347
GLM - Univariato: Opzioni 349
Funzioni aggiuntive del comando UNIANOVA 350

25 Correlazioni bivariate 351

Correlazioni bivariate: Opzioni 353
Funzioni aggiuntive dei comandi CORRELATIONS e NONPAR CORR. 353

26 Correlazioni parziali 354

Correlazioni parziali: Opzioni. 355
Opzioni aggiuntive del comando PARTIAL CORR 356

27 Distanze 357

Distanze: Misure di dissimilarità 359
Distanze: Misure di similarità 360
Opzioni aggiuntive del comando PROXIMITIES 360

28 Regressione lineare 362

Metodi di selezione della variabile di regressione lineare 363
Regressione lineare: Imposta regola. 365
Regressione lineare: grafici 365
Regressione lineare: Per salvare nuove variabili. 367
Regressione lineare: Statistiche 369
Regressione lineare: Opzioni 371
Opzioni aggiuntive del comando REGRESSION 372

29 Regressione ordinale 373

Regressione ordinale: Opzioni 374
Regressione ordinale: Output 376
Regressione ordinale: Posizione. 377
    Costruisci termini. 378
Regressione ordinale: Scala. 378
    Costruisci termini. 378
Opzioni aggiuntive del comando PLUM 379

30 Stima di curve 380

Stima di curve: Modelli 382
Stima di curve: Salva 382

31 Regressione parziale minimi quadrati 384

Modello 386
Opzioni 387

32 Analisi del vicino più vicino 389

Vicini 393
Funzioni 395
Partizioni 396
Salva 398
Output 399
Opzioni 400
Vista del modello 401
Spazio di funzioni 402
Importanza della variabile 404
Equivalenti 405
Distanze dei vicini più vicini 405
Mappa dei quadranti 406
Registro degli errori relativi alla selezione delle funzioni 407
Registro degli errori relativi alla selezione k 408
Registro degli errori relativi alla selezione k e alla selezione delle funzioni 409
Tabella di classificazione 409
Riepilogo degli errori 410

33 Analisi discriminante 411

Analisi discriminante: Definisci intervallo 413
Analisi discriminante: Seleziona casi 413
Analisi discriminante: Statistiche 414
Analisi discriminante: Metodo Stepwise 415
Analisi discriminante: Classificazione 416

Analisi discriminante: Salva 417
Funzioni aggiuntive del comando DISCRIMINANT. 418

34 Analisi fattoriale 419

Analisi fattoriale: Seleziona casi 420
Analisi fattoriale: Descrittive. 421
Analisi fattoriale: Estrazione 422
Analisi fattoriale: Rotazione 423
Analisi fattoriale: Punteggi fattoriali 424
Analisi fattoriale: Opzioni 425
Opzioni aggiuntive del comando FACTOR 426

35 Scelta di una procedura per il raggruppamento 427

36 Analisi cluster TwoStep 428

Opzioni di Analisi Cluster TwoStep 431
Grafici di Analisi Cluster TwoStep 433
Output di Analisi Cluster TwoStep 434

37 Cluster gerarchica 436

Cluster gerarchica: Metodo 438
Cluster gerarchica: Statistiche 439
Cluster gerarchica: Grafici 440
Cluster gerarchica: Salva nuove variabili 440
Funzioni aggiuntive della sintassi del comando CLUSTER 441

38 Cluster con metodo delle k-medie 442

Efficienza dell'analisi cluster K-medie 444
Cluster K-medie: Iterazioni 444
Cluster K-medie: Salva 445

Cluster K-medie: Opzioni. 445
Opzioni aggiuntive del comando QUICK CLUSTER 446

39 Test non parametrici 447

Test Chi-quadrato 447
Intervallo e valori attesi del test chi-quadrato 449
Test chi-quadrato: Opzioni 449
Funzioni aggiuntive del comando NPAR TESTS (test chi-quadrato). 450
Test binomiale 450
Test binomiale: Opzioni 451
Funzioni aggiuntive del comando NPAR TEST (test binomiale) 452
Test delle successioni. 452
Test delle successioni: Punto di divisione 453
Test delle successioni: Opzioni 453
Funzioni aggiuntive del comando NPAR TESTS (Test delle successioni) 454
Test di Kolmogorov-Smirnov per un campione. 454
Test di Kolmogorov-Smirnov per un campione: Opzioni 455
Funzioni aggiuntive del comando NPAR TESTS (Test di Kolmogorov-Smirnov per un campione) 456
Test per due campioni indipendenti. 456
Tipi di test per due campioni indipendenti 457
Test per due campioni indipendenti: Definisci gruppi 458
Test per due campioni indipendenti: Opzioni 458
Funzioni aggiuntive del comando NPAR TESTS (Due campioni indipendenti) 459
Test per due campioni dipendenti 459
Tipi di test per due campioni dipendenti 460
Test per due campioni dipendenti: Opzioni 461
Funzioni aggiuntive del comando NPAR TESTS (due campioni dipendenti) 461
Test per diversi campioni indipendenti 461
Test per diversi campioni indipendenti: tipi di test 462
Test per diversi campioni indipendenti: Definisci intervallo 463
Test per diversi campioni indipendenti: Opzioni 463
Funzioni aggiuntive del comando NPAR TESTS (K campioni indipendenti) 464
Test per diversi campioni dipendenti 464
Test per diversi campioni dipendenti: tipi di test 465
Test per diversi campioni dipendenti: Statistica 465
Funzioni aggiuntive del comando NPAR TESTS (K campioni dipendenti) 465

40 Analisi a risposta multipla 466

Risposte multiple: Definisci insiemi . . . 467
Risposte multiple: Frequenze . . . 468
Risposte multiple: Tavole di contingenza . . . 470
Risposte multiple, tavole di contingenza: Definisci intervalli delle variabili . . . 471
Risposte multiple, tavole di contingenza: Opzioni . . . 472
Funzioni aggiuntive del comando MULT RESPONSE . . . 473

41 Risultati di report 474

Report: Riepiloghi per righe . . . 474
Per ottenere un riepilogo: Riepiloghi per righe . . . 475
Formato delle colonne e di separazione del report . . . 476
Report: Linee riassuntive per/Linee riassuntive finali . . . 477
Report: Opzioni di separazione . . . 477
Report: Opzioni . . . 478
Report: Layout . . . 479
Report: Titoli . . . 480
Report: Riepiloghi per colonne . . . 480
Per ottenere un riepilogo: Riepiloghi per colonne . . . 481
Funzione di rappresentazione delle colonne di dati . . . 482
Colonna di riepilogo del totale generale . . . 483
Formato delle colonne del report . . . 484
Report: Opzioni di separazione (Riepiloghi per colonne) . . . 484
Report: Opzioni (Riepiloghi per colonne) . . . 485
Report: Layout per riepiloghi per colonne . . . 485
Funzioni aggiuntive del comando REPORT . . . 485

42 Analisi di affidabilità 487

Analisi di affidabilità: Statistiche . . . 489
Opzioni aggiuntive del comando RELIABILITY . . . 491

43 Scaling multidimensionale 492

Scaling multidimensionale: Forma dei dati . . . 494

Scaling multidimensionale: Crea misure dai dati 494
Scaling multidimensionale: Modello 495
Scaling multidimensionale: Opzioni 496
Opzioni aggiuntive del comando ALSCAL 496

44 Statistiche di rapporto 498

Statistiche di rapporto 500

45 Curva ROC 502

Curva ROC: Opzioni 504

46 Panoramica dello strumento Grafico 505

Creazione e modifica di un grafico 505
Creazione di grafici 505
Modifica dei grafici 509
Opzioni definizione grafico 512
Aggiunta e modifica di titoli e annotazioni 512
Impostazione delle opzioni generali 513

47 Strumenti 516

Informazioni sulla variabile 516
Commenti sul file di dati 517
Insiemi di variabili 517
Definisci insiemi di variabili 517
Utilizza insiemi di variabili 518
Riordinamento degli elenchi di variabili di destinazione 520

48 Opzioni 521

Opzioni generali 522

Opzioni Viewer 524
Opzioni: Dati 525
Modifica della Visualizzazione variabili predefinita 527
Opzioni: Valuta 528
Per creare formati di valuta personalizzati 529
Opzioni: Etichette di output 529
Grafici: opzioni 531
Colori degli elementi dei dati 532
Linee degli elementi dei dati 532
Simboli degli elementi dei dati 533
Riempimenti degli elementi dei dati 533
Opzioni tabella pivot 534
Opzioni di Percorsi file 535
Opzioni: Script 537
Opzioni Editor della sintassi 539
Opzioni Assegnazioni multiple 541

49 Personalizzazione dei menu e delle barre degli strumenti 543

Editor dei menu 543
Personalizzazione delle barre degli strumenti 544
Visualizza barre degli strumenti 544
Per personalizzare le barre degli strumenti 545
Proprietà barra degli strumenti 546
Modifica barra degli strumenti 546
Crea nuovo strumento 547

50 Creazione e gestione di finestre di dialogo personalizzate 549

Layout del Generatore di finestre di dialogo personalizzate 550
Creazione di una finestra di dialogo personalizzata 551
Proprietà della finestra di dialogo 551
Specifica della posizione dei menu per una finestra di dialogo personalizzata 553
Layout dei controlli sull'area di disegno 554
Creazione del modello di sintassi 555
Visualizzazione dell'anteprima di una finestra di dialogo personalizzata 557
Gestione delle finestre di dialogo personalizzate 558
Tipi di controlli 560

Elenco origine . . . 561
Lista destinazione . . . 561
Applicazione di filtri agli elenchi di variabili . . . 563
Casella di controllo . . . 563
Casella combinata . . . 564
Controllo testo . . . 565
Controllo numerico . . . 566
Controllo testo statico . . . 566
Gruppo di elementi . . . 567
Gruppo pulsanti di opzione . . . 567
Gruppo di caselle di controllo . . . 568
Browser di file . . . 569
Pulsante sotto-finestra . . . 571
Finestre di dialogo personalizzate per i comandi di estensione . . . 572
Creazione di versioni localizzate di finestre di dialogo personalizzate . . . 574

51 Sessioni Utilità SPSS Production *576*

Opzioni HTML . . . 579
Opzioni di PowerPoint . . . 580
Opzioni PDF . . . 580
Opzioni testo . . . 580
Valori di runtime . . . 580
Prompt utente . . . 582
Esecuzione di sessioni dell'utilità SPSS Production dalla riga di comando . . . 583
Conversione dei file della sessione Utilità SPSS Production . . . 584

52 Sistema di gestione dell'output *585*

Tipi di oggetto di output . . . 587
Identificatori di comando e sottotipi di tabelle . . . 589
Etichette . . . 590
Opzioni SGO . . . 591
Registrazione . . . 596
Esclusione della visualizzazione dell'output in Viewer . . . 597
Trasmissione dell'output ai file di dati in formato SPSS Statistics . . . 597
Esempio: tabella bidimensionale singola . . . 597
Esempio: Tabelle con strati . . . 598
File dati creati con tabelle multiple . . . 599

Uso degli elementi colonna per il controllo delle variabili nei file di dati 602
Nomi di variabili nei file dati generati da SGO 604
Struttura della tabella OXML 605
Identificatori SGO 608
Copia degli identificatori SGO dalla vista riassuntiva del Viewer 610

Appendice

A Convertitore di sintassi dei comandi TABLES e IGRAPH 612

Indice 615

Capitolo
1

Osservazioni generali

SPSS Statistics è un valido sistema per l'analisi statistica e la gestione dei dati in un ambiente grafico in cui è possibile utilizzare menu descrittivi e semplici finestre di dialogo per eseguire automaticamente numerose operazioni. Con un solo clic del mouse è possibile eseguire la maggior parte delle operazioni.

Oltre all'interfaccia per l'analisi statistica, di facile utilizzo in quanto prevede l'impiego del mouse per numerose operazioni, in SPSS Statistics sono disponibili anche:

Editor dei dati. L'Editor dei dati è un sistema versatile dotato delle stesse caratteristiche di un foglio elettronico per la definizione, l'inserimento, la modifica e la visualizzazione dei dati.

Viewer. Il Viewer consente di sfogliare rapidamente i risultati, visualizzare e nascondere l'output in modo selettivo, modificare l'ordine di visualizzazione dei risultati e trasferire tabelle e grafici di alta qualità in o da altre applicazioni.

Tabelle pivot multidimensionali. È possibile mettere in risalto i risultati grazie alle tabelle pivot multidimensionali. In queste tabelle i risultati possono essere visualizzati in modi diversi modificando la disposizione di righe, colonne e strati. È possibile mettere in risalto risultati importanti che in genere non compaiono nei tipi di report più comuni. I gruppi possono essere facilmente confrontati dividendo la tabella in modo da visualizzare un solo gruppo per volta.

Grafici ad alta risoluzione. Sono inclusi come funzioni standard grafici a torta, grafici a barre, istogrammi, grafici a dispersione, grafici 3-D e di altro tipo, tutti a colori e ad alta risoluzione.

Accesso ai database. È possibile recuperare informazioni dai database utilizzando Creazione guidata di query anzichè query SQL complicate.

Trasformazioni di dati. Le funzioni di trasformazione risultano utili per preparare i dati alle analisi. È possibile creare sottoinsiemi di dati, unire categorie, aggiungere, aggregare, unire, dividere e trasporre i file ed eseguire molte altre operazioni.

Aiuto in linea. Le esercitazioni dettagliate forniscono una panoramica completa del programma; l'Aiuto sensibile al contesto disponibile in tutte le finestre di dialogo consente di visualizzare informazioni relative a operazioni specifiche; le definizioni popup dei risultati delle tabelle pivot aiutano a comprendere i termini statistici; le Analisi guidate consentono di trovare rapidamente le procedure necessarie e gli Studi di casi offrono esempi pratici sull'utilizzo delle procedure statistiche e costituiscono un valido supporto all'interpretazione dei risultati.

Linguaggio a comandi. Benché la maggior parte delle attività possa essere eseguita con un semplice clic del mouse, in SPSS Statistics è disponibile anche un linguaggio a comandi che consente di salvare e automatizzare molte delle operazioni più comuni. Tale linguaggio offre inoltre alcune funzionalità a cui non è possibile accedere tramite i menu e le finestre di dialogo.

La documentazione completa sulla sintassi dei comandi è integrata nella Guida in linea ed è disponibile su un documento PDF a parte, *Command Syntax Reference*, anch'esso richiamabile dal menu ?.

Novità della versione 17.0

Nuovo editor della sintassi. L'editor della sintassi è stato interamente ridisegnato con funzioni quali l'autocompletamento, la codifica colori, i segnalibri e i punti di interruzione. L'autocompletamento fornisce un elenco di nomi di comandi, sottocomandi e parole chiave validi, consentendo di risparmiare tempo quando si fa riferimento ai grafici della sintassi. La codifica colori consente di rilevare rapidamente i termini non riconosciuti, come ad esempio alcuni errori di sintassi tipici. I segnalibri permettono di navigare rapidamente all'interno di grandi file di sintassi dei comandi. I punti di interruzione consentono di interrompere l'esecuzione in un punto specifico, in modo da ispezionare i dati o l'output prima di procedere. Per ulteriori informazioni, vedere Utilizzo dell'Editor della sintassi in Capitolo 13 a pag. 272.

Generatore di finestre di dialogo personalizzate. Il Generatore di finestre di dialogo personalizzate consente di creare e gestire le finestre di dialogo personalizzate per la generazione della sintassi dei comandi. È possibile creare finestre di dialogo personalizzate per generare la sintassi di più comandi, inclusi i comandi di estensione personalizzati implementanti in Python o R. Per ulteriori informazioni, vedere Creazione e gestione di finestre di dialogo personalizzate in Capitolo 50 a pag. 549.

Supporto di più lingue. Oltre alla possibilità di cambiare la lingua dell'output, funzione disponibile già nelle versioni precedenti, ora è possibile anche modificare la lingua dell'interfaccia utente. Per ulteriori informazioni, vedere Opzioni generali in Capitolo 48 a pag. 522.

Informazioni sui dati. La procedura Informazioni sui dati restituisce informazioni del dizionario, ad esempio nomi di variabili, etichette di variabili e di valori o valori mancanti, e statistiche riassuntive per alcune o tutte le variabili e gli insiemi a risposta multipla presenti nell'insieme di dati attivo. Per le variabili nominali e ordinali e gli insiemi a risposta multipla, le statistiche riassuntive includono conteggi e percentuali. Per le variabili di scala, le statistiche riassuntive includono media, deviazione standard e quartili. Per ulteriori informazioni, vedere Informazioni sui dati in Capitolo 14 a pag. 282.

Analisi del vicino più vicino. L'analisi del vicino più vicino è un metodo per la classificazione dei casi basato sulla similarità ad altri casi. Nell'apprendimento automatico, questo metodo è stato sviluppato per riconoscere modelli di dati senza richiedere una corrispondenza esatta con eventuali modelli o casi archiviati. I casi simili sono vicini gli uni agli altri, mentre i casi dissimili sono distanti gli uni dagli altri. Pertanto, la distanza tra due casi rappresenta una misura della loro dissimilarità. Per ulteriori informazioni, vedere Analisi del vicino più vicino in Capitolo 32 a pag. 389.

Assegnazione multipla. La procedura di Assegnazione multipla esegue più assegnazioni dei valori di dati mancanti. Dato un insieme di dati contenente dei valori mancanti, tale procedura genera uno o più insiemi di dati in cui i valori mancanti vengono sostituiti da stime plausibili. Successivamente è possibile ottenere dei risultati raggruppati quando si eseguono altre procedure.

La procedura riepiloga inoltre i valori mancanti nell'insieme di dati di lavoro. Questa funzione è disponibile nel modulo aggiuntivo Missing Values.

Analisi RFM. L'analisi RFM (recency, frequency, monetary) è una tecnica utilizzata per identificare i clienti esistenti che più probabilmente risponderanno a una nuova offerta. Questa tecnica è comunemente usata nel direct marketing. Questa funzione è disponibile nel modulo aggiuntivo EZ RFM.

Miglioramenti in Regressione categoriale. La funzione di Regressione categoriale è stata migliorata in modo da includere i metodi di regolarizzazione e campionamento ripetuto per valutare e ottimizzare la precisione predittiva. Insieme, questi nuovi metodi rendono possibile la creazione di modelli allo stato dell'arte, anche per grandi volumi di dati (in cui il numero di variabili è maggiore delle osservazioni, come ad esempio nella genomica). Questa funzione è disponibile nell'opzione aggiuntiva Categorie.

Lavagna grafica. Le visualizzazioni Lavagna grafica includono grafici e diagrammi creati da un modello di visualizzazione. SPSS Statistics viene fornito con modelli di visualizzazione incorporati. È inoltre possibile utilizzare un prodotto distinto, SPSS Viz Designer, per creare i propri modelli di visualizzazione. I nuovi modelli di visualizzazione sono effettivamente tipi di visualizzazione personalizzata.

Esportazione dell'output. Più opzioni di formato per l'esportazione dell'output e più controllo sui contenuti esportati, tra cui:

- Andare a capo o ridurre tabelle di grandi dimensioni nei documenti Word. Per ulteriori informazioni, vedere Opzioni Word/RTF in Capitolo 10 a pag. 226.
- Creare nuovi fogli di lavoro o accodare i dati a fogli di lavoro esistenti in una cartella di lavoro Excel. Per ulteriori informazioni, vedere Opzioni Excel in Capitolo 10 a pag. 227.
- Salvare le specifiche dell'esportazione dell'output sotto forma di sintassi di comandi con il comando `OUTPUT EXPORT`. Tutte le funzioni per l'esportazione dell'output nella finestra di dialogo Esporta output sono ora disponibili in una sintassi di comando; in questo modo è possibile salvare e rieseguire le specifiche dell'esportazione e includerle in sessioni di utilità di SPSS Production automatizzate.
- Ora il Sistema di gestione dell'output (SGO) supporta questi formati di output aggiuntivi: Word, Excel e PDF. Per ulteriori informazioni, vedere Sistema di gestione dell'output in Capitolo 52 a pag. 585.

Sposta valori. Sposta valori crea nuove variabili contenenti i valori di variabili esistenti in casi precedenti (ritardo) o successivi (anticipo). Per ulteriori informazioni, vedere Sposta valori in Capitolo 8 a pag. 142.

Miglioramenti dell'aggregazione. Ora è possibile utilizzare le funzioni della procedura Aggrega senza specificare una variabile di separazione. Per ulteriori informazioni, vedere Aggrega dati in Capitolo 9 a pag. 188.

Funzione mediana. Ora è disponibile una funzione mediana per calcolare il valore mediano in diverse variabili per ciascun caso.

Finestre

In SPSS Statistics sono disponibili diversi tipi di finestre:

Editor dei dati. Nell'Editor dei dati viene visualizzato il contenuto del file di dati. Nell'Editor dei dati è possibile creare nuovi file di dati o modificare quelli esistenti. Se ci sono più file di dati aperti, ciascun file di dati viene visualizzato una finestra a parte dell'Editor dei dati.

Viewer. Tutti i risultati statistici, le tabelle e i grafici vengono visualizzati nel Viewer. È possibile modificare l'output e salvarlo per un successivo utilizzo. Una finestra del Viewer viene aperta automaticamente la prima volta che viene eseguita una procedura che genera output.

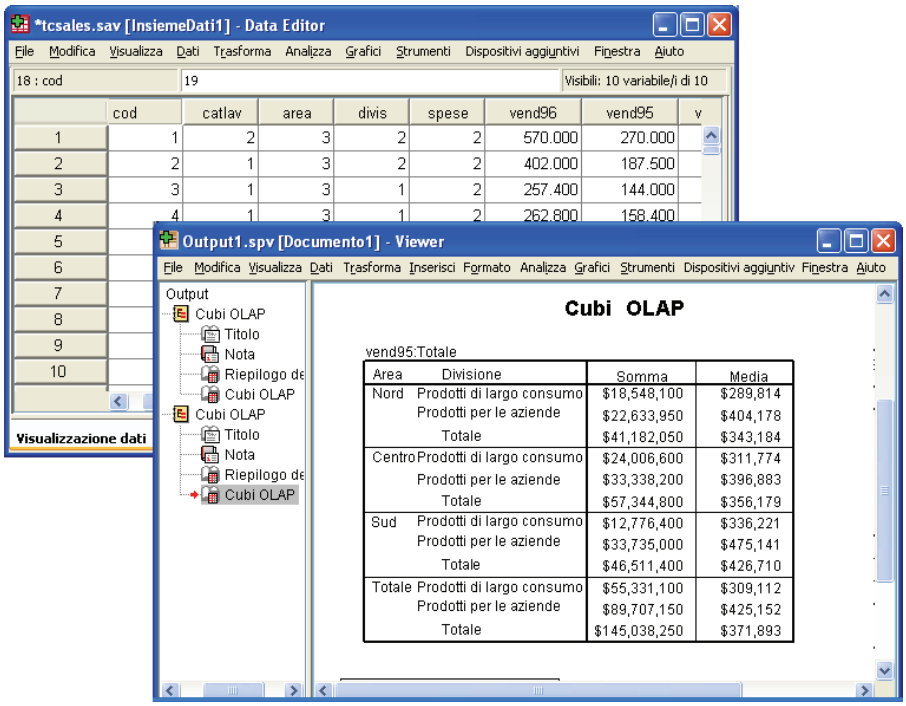
Editor di tabelle pivot. L'output visualizzato nelle tabelle pivot può essere modificato in molti modi nell'Editor di tabelle pivot. È possibile modificare il testo, scambiare i dati contenuti in righe e colonne, aggiungere colore, creare tabelle multidimensionali e visualizzare o nascondere i risultati in modo selettivo.

Editor dei grafici. È possibile modificare grafici ad alta risoluzione nelle finestre dei grafici. È possibile modificare i colori, selezionare diversi tipi o dimensioni di carattere, scambiare gli assi orizzontale e verticale, ruotare grafici a dispersione 3-D e perfino cambiare tipo di grafico.

Editor di output testuale. L'output testuale non visualizzato nelle tabelle pivot può essere modificato nell'Editor di output testuale. È possibile modificare l'output e modificare le caratteristiche del carattere (tipo, stile, colore, dimensioni).

Editor della sintassi. È possibile incollare le impostazioni della finestra di dialogo in una finestra della sintassi quando tali impostazioni vengono visualizzate sotto forma di sintassi. È possibile modificare la sintassi dei comandi per utilizzare particolari funzioni che non sono disponibili nelle finestre di dialogo. È possibile salvare tali comandi in un file da utilizzare nelle sessioni successive.

Figura 1-1
Editor dei dati e Viewer


*tcsales.sav [InsiemeDati1] - Data Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntivi Finestra Aiuto
18 : cod
19
Visibili: 10 variabile/i di 10
cod catlav area divis spese vend96 vend95
1 1 2 3 2 2 570.000 270.000
2 2 1 3 2 2 402.000 187.500
3 3 1 3 1 2 257.400 144.000
4 4 1 3 1 2 262.800 158.400
Visualizzazione dati
Output1.spv [Documento1] - Viewer
File Modifica Visualizza Dati Trasforma Inserisci Formato Analizza Grafici Strumenti Dispositivi aggiuntiv Finestra Aiuto
Output
Cubi OLAP
Titolo
Nota
Riepilogo de
Cubi OLAP
Cubi OLAP
Titolo
Nota
Riepilogo de
Cubi OLAP
Cubi OLAP
vend95:Totale
Area Divisione Somma Media
Nord Prodotti di largo consumo $18,548,100 $289,814
Prodotti per le aziende $22,633,950 $404,178
Totale $41,182,050 $343,184
Centro Prodotti di largo consumo $24,006,600 $311,774
Prodotti per le aziende $33,338,200 $396,883
Totale $57,344,800 $356,179
Sud Prodotti di largo consumo $12,776,400 $336,221
Prodotti per le aziende $33,735,000 $475,141
Totale $46,511,400 $426,710
Totale Prodotti di largo consumo $55,331,100 $309,112
Prodotti per le aziende $89,707,150 $425,152
Totale $145,038,250 $371,893

Confronto tra finestra designata e attiva

Se sono aperte più finestre del Viewer, l'output viene trasferito nella finestra **designata**. Se sono aperte più finestre dell'Editor della sintassi, la sintassi viene incollata nella finestra designata. Le finestre designate sono contrassegnate riportano un simbolo più nell'icona della barra del titolo. È possibile cambiare le finestre designate in qualsiasi momento.

La finestra designata non deve essere confusa con la finestra **attiva**, che è la finestra correntemente selezionata. Se le finestre sono sovrapposte, la finestra attiva viene visualizzata in primo piano. Se si apre una nuova finestra, questa diventerà automaticamente la finestra attiva e la finestra designata.

Per cambiare la finestra designata del Viewer o dell'Editor della sintassi

- Fare clic in un punto qualsiasi della finestra per renderla attiva.
- Fare clic sul pulsante Designa finestra sulla barra degli strumenti (l'icona con il simbolo più).

  o

- Dai menu, scegliere:

  Strumenti
  Designa finestra

Nota: per le finestre dell'Editor dei dati, la finestra attiva determina il file di dati che verrà usato per i calcoli o le analisi successivi. In questo caso non è disponibile nessuna finestra "designata". Per ulteriori informazioni, vedere Gestione di più sorgenti di dati in Capitolo 6 a pag. 103.

Barra di stato

Nella barra di stato, situata nella parte inferiore di ciascuna finestra di SPSS Statistics, vengono visualizzate le seguenti informazioni:

Stato del file dati dei comandi. Per ogni procedura o comando eseguito, verrà indicato il numero di casi già elaborati. Per le procedure statistiche che richiedono un'elaborazione iterativa, verrà visualizzato il numero di iterazioni.

Stato filtro. Se è stato selezionato un campione casuale o un sottoinsieme di casi per l'analisi, il messaggio Filtro attivo indica che è attivo un filtro e che non tutti i casi del file di dati verranno inclusi nell'analisi.

Stato della ponderazione. Il messaggio Peso attivo indica che per pesare i casi dell'analisi viene utilizzata una variabile di ponderazione.

Stato Dividi. Il messaggio Dividi attivo indica che il file di dati è stato suddiviso in gruppi distinti ai fini dell'analisi in base ai valori di una o più variabili di raggruppamento.

Finestre di dialogo

La maggior parte dei comandi dei menu di SPSS consente di aprire finestre di dialogo, dove è possibile selezionare variabili e opzioni per l'analisi.

Le finestre di dialogo delle procedure statistiche e dei grafici hanno in genere due componenti di base:

Elenco di variabili sorgente. Elenco delle variabili del file di dati attivo. dove vengono visualizzati solo i tipi di variabile consentiti dalla procedura selezionata. L'utilizzo delle variabili di stringa corta e lunga è limitato in molte procedure.

Elenco di variabili di destinazione. Le variabili scelte per l'analisi vengono visualizzate in una o più liste, ad esempio nelle liste delle variabili dipendenti e indipendenti.

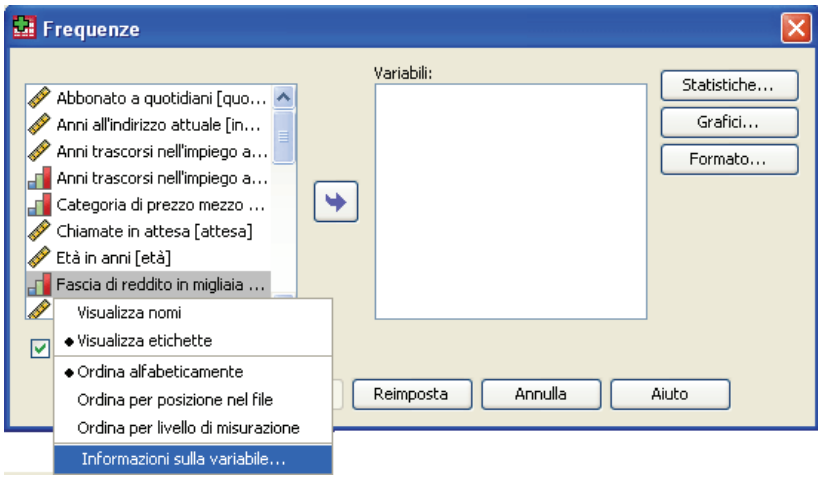
Nomi ed etichette di variabili negli elenchi delle finestre di dialogo

Negli elenchi delle finestre di dialogo è possibile visualizzare i nomi o le etichette delle variabili e negli elenchi delle variabili sorgente è possibile controllare l'ordinamento delle variabili.

- Per controllare gli attributi di visualizzazione predefiniti delle variabili negli elenchi sorgente, scegliere Opzioni dal menu Modifica. Per ulteriori informazioni, vedere Opzioni generali in Capitolo 48 a pag. 522.
- Per modificare gli attributi di visualizzazione degli elenchi delle variabili sorgente all'interno di una finestra di dialogo, fare clic con il pulsante destro del mouse su una variabile qualsiasi nell'elenco sorgente e selezionare gli attributi di visualizzazione dal menu di scelta rapida. È possibile visualizzare i nomi o le etichette delle variabili (per ogni variabile i nomi vengono

visualizzati senza etichette definite) e l'elenco sorgente può essere ordinato in base all'ordine del file, all'ordine alfabetico o in base al livello di misurazione. Per ulteriori informazioni sul livello di misurazione, vedere Icone relative al tipo di dati, al livello di misurazione e all'elenco delle variabili a pag. 8.

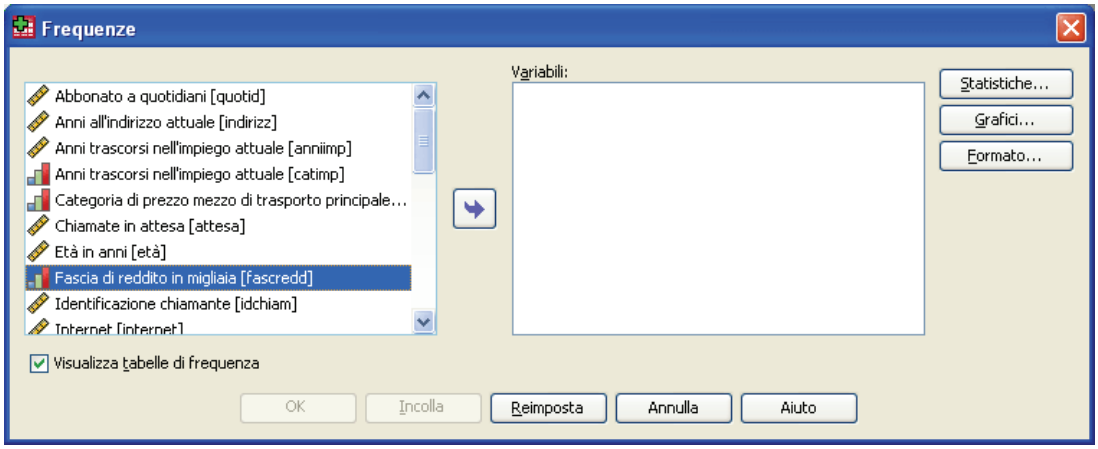
Figura 1-2
Etichette di variabili visualizzate in una finestra di dialogo



Ridimensionamento delle finestre di dialogo

È possibile ridimensionare le finestre di dialogo proprio come per le finestre, selezionando e trascinando i bordi o gli angoli esterni. Ad esempio, se la finestra di dialogo viene allargata, anche gli elenchi delle variabili risulteranno più ampi.

Figura 1-3
Finestra di dialogo ridimensionata



Controlli delle finestre di dialogo

Nella maggior parte delle finestre di dialogo di SPSS sono disponibili cinque controlli standard:

OK. Consente di eseguire la procedura. Dopo aver selezionato le variabili e impostato le eventuali opzioni aggiuntive, fare clic su OK per eseguire la procedura e chiudere la finestra di dialogo.

Incolla. Consente di generare la sintassi in base alle impostazioni della finestra di dialogo e di incollarla in una finestra della sintassi. È possibile personalizzare i comandi utilizzando funzioni aggiuntive di SPSS, non disponibili nelle finestre di dialogo.

Ripristina. Consente di deselezionare qualsiasi variabile nell'elenco delle variabili selezionate e di ripristinare tutte le impostazioni originali della finestra di dialogo e delle eventuali sottofinestre.

Annulla. Consente di annullare qualsiasi modifica apportata alle impostazioni della finestra di dialogo a partire dall'ultima apertura e quindi di chiudere la finestra stessa. Le impostazioni delle finestre di dialogo vengono mantenute per l'intera sessione attiva di SPSS. Le ultime impostazioni specificate nella finestra di dialogo verranno mantenute fino alla modifica successiva.

Aiuto. Visualizza l'aiuto contestuale. Questo controllo consente di accedere a una finestra dell'Aiuto in cui sono visualizzate informazioni relative alla finestra di dialogo corrente.

Selezione di variabili

Per selezionare una singola variabile, è sufficiente selezionarla nell'elenco delle variabili sorgente e trascinarla nell'elenco delle variabili di destinazione. È anche possibile utilizzare il pulsante freccia per spostare le variabili dall'elenco sorgente agli elenchi di destinazione. Se esiste un solo elenco di destinazione delle variabili, è possibile fare clic sulle singole variabili per spostarle dall'elenco sorgente a quello di destinazione.

È inoltre possibile selezionare più variabili:

- Per selezionare più variabili raggruppate nell'elenco di variabili, fare clic sulla prima variabile, quindi, tenendo premuto Maiusc, fare clic sull'ultima variabile del gruppo.
- Per selezionare più variabili non raggruppate nell'elenco delle variabili, fare clic sulla prima variabile, quindi premere Ctrl e fare clic sulla variabile successiva e così via (in Macintosh: tenere premuto Comando e fare clic).

Icone relative al tipo di dati, al livello di misurazione e all'elenco delle variabili

Le icone visualizzate accanto alla variabili negli elenchi delle finestre di dialogo forniscono informazioni sul tipo di variabile e sul livello di misurazione.

Livello di misurazione	**Tipo di dati**			
	Numerica	Stringa	Data	Ora
Scala		n/a		

Ordinale				
Nominale				

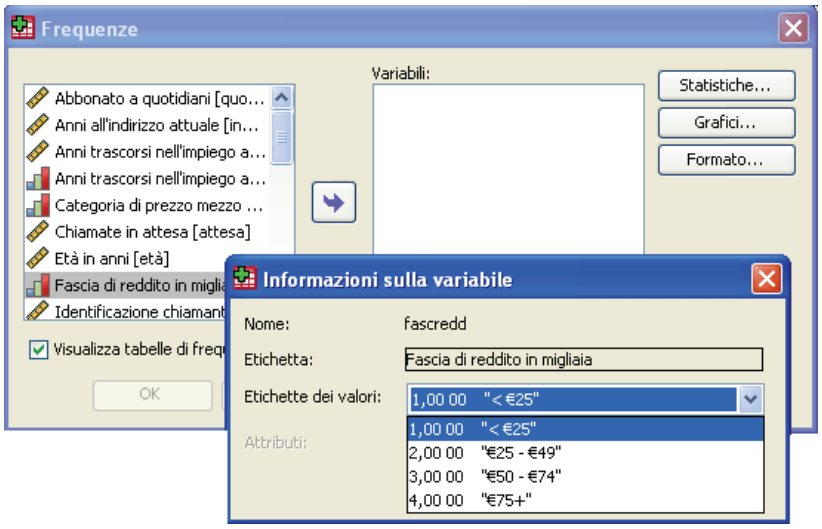
- Per ulteriori informazioni sul livello di misurazione, vedere Livello di misurazione delle variabili a pag. 80.
- Per ulteriori informazioni sui tipi di dati numerici, di stringa, di data e temporali, vedere Tipo di variabile a pag. 81.

Per ottenere informazioni sulle variabili in una finestra di dialogo

▶ Fare clic con il pulsante destro del mouse su una variabile nell'elenco di variabili sorgente o di destinazione.

▶ Selezionare Informazioni variabili.

Figura 1-4
Informazioni sulle variabili



Passaggi fondamentali dell'analisi dei dati

In SPSS Statistics è facile analizzare i dati. È sufficiente:

Inserire i dati in SPSS Statistics. È possibile aprire un file di dati di SPSS Statistics precedentemente salvato, leggere un foglio elettronico, un database o un file dati di testo oppure inserire i dati direttamente nell'Editor dei dati.

Selezionare una procedura. Selezionare una procedura dai menu per calcolare statistiche o creare un grafico.

Selezionare le variabili per l'analisi. Le variabili del file di dati verranno visualizzate in una finestra di dialogo relativa alla procedura.

Eseguire la procedura e verificare i risultati. I risultati vengono visualizzati nel Viewer.

Statistics Coach

Se non si ha familiarità con SPSS Statistics o con le procedure statistiche disponibili, è consigliabile consultare Statistics Coach, che offre un rapido metodo di apprendimento basato su domande semplici formulate in un linguaggio comune con esempi visivi che permettono di individuare le funzioni statistiche e per la creazione di grafici più adatte ai dati.

Per utilizzare Statistics Coach, dai menu di qualsiasi finestra di SPSS Statistics scegliere:
?
Statistics Coach

In Statistics Coach vengono illustrate solo alcune procedure del sistema Base. Viene fornita assistenza generica in relazione a molte delle tecniche statistiche di maggiore utilizzo.

Ulteriori informazioni

Per informazioni generali sugli elementi di base, vedere l'esercitazione in linea. Da qualsiasi barra dei menu di SPSS Statistics scegliere:
?
Esercitazione

Capitolo
2

Aiuto

L'assistenza per l'uso del programma viene fornita in molti modi:

Menu ?. Il menu Aiuto visualizzato nella maggior parte delle finestre consente di accedere alla guida principale, alle esercitazioni e al materiale tecnico di riferimento.

- **Argomenti.** Consente di accedere alle schede Sommario, Indice e Cerca da utilizzare per trovare specifici argomenti dell'Aiuto.
- **Esercitazione.** Istruzioni guidate illustrate che spiegano come usare molte delle funzioni di base. Non è necessario visualizzare tutta l'esercitazione dall'inizio alla fine. È possibile infatti scegliere gli argomenti che si desiderano visualizzare, selezionare gli argomenti desiderati in qualsiasi ordine ed utilizzare l'indice o il sommario per trovare argomenti specifici.
- **Studi di casi.** Esempi pratici che spiegano come creare vari tipi di analisi statistiche ed interpretarne i risultati. Sono inoltre disponibili i file di dati utilizzati negli esempi in modo da seguire questi ultimi e capire come sono stati ottenuti i risultati prodotti. È possibile scegliere le procedure specifiche sulle quali si desiderano maggiori informazioni dal sommario oppure ricercare gli argomenti desiderati nell'indice.
- **Statistics Coach.** Analisi di tipo guidato che consentono di trovare la procedura corretta. Appena si selezionano gli argomenti desiderati, le Analisi guidate aprono la finestra di dialogo corrispondente alle procedure per la creazione di statistiche, report o grafici che soddisfano i criteri selezionati. La funzione Analisi guidate consente di accedere alla maggior parte delle procedure statistiche e di reporting del modulo Base nonché a molte delle procedure per la creazione di grafici.
- **Sintassi.** Informazioni dettagliate sui riferimenti alla sintassi dei comandi sono disponibili in due formati: disponibili nella Guida integrata e su un documento PDF a parte in *Command Syntax Reference*, anch'esso consultabile tramite il menu ?.
- **Algoritmi statistici.** Gli algoritmi usati per la maggior parte delle procedure statistiche sono disponibili in due formati: nella Guida in linea e in un documento PDF distinto disponibile nel CD dei manuali. Per collegamenti ad algoritmi specifici nella Guida in linea, selezionare Algoritmi nel menu ?.

Aiuto contestuale. L'Aiuto contestuale può essere richiamato da più punti dell'interfaccia utente.

- **Pulsanti Aiuto nelle finestre di dialogo.** Nella maggior parte delle finestre di dialogo è disponibile un pulsante Aiuto che consente di passare direttamente a un argomento dell'Aiuto relativo a quella finestra di dialogo. Nell'argomento dell'Aiuto vengono fornite informazioni generali e collegamenti agli argomenti correlati.

- **Aiuto contestuale nelle tabelle pivot.** Fare clic con il pulsante destro del mouse su un termine di una tabella pivot attiva nel Viewer e scegliere Guida rapida dal menu di scelta rapida per visualizzare la relativa definizione.
- **Sintassi dei comandi.** Nella finestra di una sintassi di comandi, spostare il cursore su qualsiasi punto del blocco della sintassi relativa ad un comando e premere F1 sulla tastiera. Viene visualizzato un grafico completo della sintassi dei comandi per il comando selezionato. La documentazione completa sulla sintassi dei comandi può essere visualizzata selezionando i collegamenti nell'elenco di argomenti correlati o la scheda Contenuti dell'Aiuto.

Altre risorse

Sito Web del Supporto tecnico. Per trovare una risposta a molti problemi comuni, è possibile visitare il sito Web *http://support.spss.com*. Per accedere al sito Web del supporto tecnico, sono necessari un ID di accesso e una password. Per informazioni su come richiedere l'ID e la password, consultare l'URL sopra specificato.

Developer Central. Developer Central contiene risorse per utenti e sviluppatori di applicazioni di qualsiasi livello. È possibile scaricare utilità, esempi di grafici, nuovi moduli statistici e articoli. È possibile visitare il sito Web Developer Central all'indirizzo *http://www.spss.com/devcentral*.

Visualizzazione di informazioni sui termini delle tabelle pivot dell'output

Per visualizzare la definizione relativa a un termine nell'output di una tabella pivot nel Viewer:

- ▶ Fare doppio clic nella tabella pivot per attivarla.
- ▶ Con il pulsante destro del mouse fare clic sul termine per il quale si desidera visualizzare una descrizione.
- ▶ Scegliere Guida rapida dal menu di scelta rapida.

La definizione del termine verrà visualizzata in una finestra popup.

Figura 2-1
Visualizzazione dell'Aiuto per il glossario di una tabella pivot attivata tramite il pulsante destro del mouse

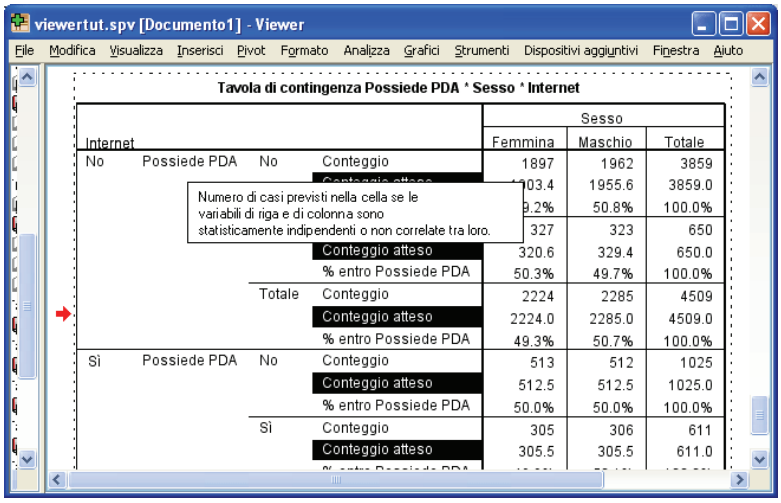

viewertut.spv [Documento1] - Viewer
File Modifica Visualizza Inserisci Pivot Formato Analizza Grafici Strumenti Dispositivi aggiuntivi Finestra Aiuto
Tavola di contingenza Possiede PDA * Sesso * Internet
Numero di casi previsti nella cella se le variabili di riga e di colonna sono statisticamente indipendenti o non correlate tra loro.

Tavola di contingenza Possiede PDA * Sesso * Internet

Internet				Sesso		
				Femmina	Maschio	Totale
No	Possiede PDA	No	Conteggio	1897	1962	3859
			Conteggio atteso	1903.4	1955.6	3859.0
			% entro Possiede PDA	49.2%	50.8%	100.0%
		Sì	Conteggio	327	323	650
			Conteggio atteso	320.6	329.4	650.0
			% entro Possiede PDA	50.3%	49.7%	100.0%
		Totale	Conteggio	2224	2285	4509
			Conteggio atteso	2224.0	2285.0	4509.0
			% entro Possiede PDA	49.3%	50.7%	100.0%
Sì	Possiede PDA	No	Conteggio	513	512	1025
			Conteggio atteso	512.5	512.5	1025.0
			% entro Possiede PDA	50.0%	50.0%	100.0%
		Sì	Conteggio	305	306	611
			Conteggio atteso	305.5	305.5	611.0

Capitolo
3

File di dati

I file di dati sono disponibili in numerosi formati, molti dei quali possono essere gestiti in SPSS, ad esempio:

- Fogli elettronici creati con Excel e Lotus
- Le tabelle dei database di più sorgenti, compresi Oracle, SQLServer, Access, dBASE e altri
- File di testo delimitati da tabulazioni e altri tipi di file di testo
- File di dati in formato SPSS Statistics creati in altri sistemi operativi.
- File di dati SYSTAT.
- File di dati SAS
- File di dati Stata

Apertura di file di dati

Oltre ai file salvati in formato SPSS Statistics, è possibile aprire file di Excel, SAS, Stata, file delimitati da tabulazioni e altri file senza doverli convertire in un formato intermedio né inserire informazioni di definizione dei dati.

- L'apertura di un file di dati comporta l'attivazione dell'insieme di dati. Se ci sono altri file di dati aperti, questi rimangono aperti e possono quindi essere usati durante la sessione. Fare clic in un qualsiasi punto della finestra dell'Editor dei dati per impostare qualsiasi file di dati aperto come attivo. Per ulteriori informazioni, vedere Utilizzo di più sorgenti dati in Capitolo 6 a pag. 102.
- In modalità analisi distribuita con un server remoto per l'elaborazione dei comandi e l'esecuzione delle procedure, i file di dati, le cartelle e le unità disponibili dipendono da cosa viene messo a disposizione dal server remoto. Il nome del server corrente è indicato nella parte superiore della finestra di dialogo. Non è possibile accedere a file di dati sul computer locale a meno che l'unità venga specificata come dispositivo condiviso e le cartelle contenenti quei file come cartelle condivise. Per ulteriori informazioni, vedere Analisi distribuita in Capitolo 4 a pag. 69.

Per aprire file di dati

► Dai menu, scegliere:

File
  Apri
    Dati...

► Nella finestra di dialogo Apri dati, selezionare il file che si desidera aprire.

▶ Fare clic su Apri.

Se si desidera, è possibile:

- Impostare automaticamente la larghezza di ogni variabile di stringa sul valore più lungo osservato per tale variabile utilizzando l'opzione Minimizza le larghezze delle stringhe sulla base dei valori osservati. Ciò si rivela particolarmente utile al momento della lettura di file di dati di pagina di codice in modalità Unicode. Per ulteriori informazioni, vedere Opzioni generali in Capitolo 48 a pag. 522.
- Leggere i nomi delle variabili dalla prima riga di file di fogli elettronici.
- Specificare un intervallo di celle da leggere dai fogli elettronici.
- Specificare un foglio di lavoro del file di Excel da leggere (Excel 95 o versione successiva).

Per informazioni sulla lettura di dati da database, vedere Lettura di file di database a pag. 17. Per informazioni sulla lettura di dati da file di dati testo, vedere Importazione guidata di testo a pag. 32.

Tipi di file di dati

SPSS Statistics. Consente di aprire file di dati salvati in formato SPSS Statistics e anche il prodotto SPSS/PC+ per DOS.

SPSS/PC+. Consente di aprire file di dati SPSS/PC+.

SYSTAT. Consente di aprire file di dati SYSTAT.

SPSS Statistics Portable. Consente di aprire file di dati salvati in formato portabile. Il salvataggio di file in formato portabile richiede tempi notevolmente più lunghi del salvataggio in formato SPSS Statistics.

Excel. Consente di aprire i file di Excel.

Lotus 1-2-3. Consente di aprire i file salvati in formato 1-2-3 per le versioni 3.0, 2.0 o 1A di Lotus

SYLK. Consente di aprire file di dati salvati in formato SYLK (Symbolic Link), utilizzato da molte applicazioni per fogli elettronici.

dBASE. Consente di aprire file in formato dBASE per dBASE IV, dBASE III, III PLUS o dBASE II. Ciascun caso rappresenta un record. Quando si salva un file in questo formato, le etichette delle variabili e dei valori e le specificazioni dei casi mancanti andranno perdute.

SAS. Versioni SAS 6–9 e file di trasporto SAS.

Stata. Versioni Stata 4–8.

Apertura file: opzioni

Leggi nomi di variabile. Per i fogli elettronici, è possibile leggere i nomi delle variabili dalla prima riga del file o dalla prima riga dell'intervallo specificato. I valori vengono convertiti in base alle necessità per creare nomi di variabili validi, inclusa la conversione degli spazi in caratteri di sottolineatura.

Foglio di lavoro. I file di Excel 95 o versioni successive possono contenere più fogli di lavoro. Per impostazione predefinita, l'Editor dei dati legge il primo foglio di lavoro. Per leggere un altro foglio di lavoro, selezionare il foglio desiderato dall'elenco a discesa.

Intervallo. Per i file di dati di fogli elettronici è inoltre possibile leggere un intervallo di celle. Per specificare gli intervalli di celle è possibile utilizzare lo stesso metodo adottato nell'applicazione per fogli elettronici.

Lettura di file di Excel 95 o versione successiva

Per la lettura dei file di Excel 95 o versione successiva sono valide le seguenti regole:

Tipo e larghezza dei dati. Ciascuna colonna rappresenta una variabile. Il tipo e la larghezza dei dati di ciascuna variabile vengono determinati dal tipo e dalla larghezza dei dati del file di Excel. Se nella colonna sono presenti più tipi di dati (ad esempio, data e numerici), verrà impostato il tipo di dati stringa e tutti i valori verranno letti come valori stringa validi.

Celle vuote. Per le variabili numeriche, le celle vuote vengono convertite al valore mancante di sistema, indicato da un punto. Per le variabili stringa, uno spazio rappresenta un valore stringa valido. Le celle vuote, pertanto, vengono considerate come valori stringa validi.

Nomi delle variabili. Se la prima riga del file di Excel o la prima riga dell'intervallo specificato vengono lette come nomi delle variabili, i nomi non conformi alle regole di denominazione delle variabili verranno convertiti in nomi di variabili validi e i nomi originali verranno utilizzati come etichette delle variabili. Se i nomi delle variabili non vengono letti dal file di Excel, verranno assegnati i nomi di variabili predefiniti.

Lettura di file Excel di versioni precedenti e di altri fogli elettronici

Le regole riportate di seguito sono valide per la lettura di file di Excel antecedenti a Excel 95 e di file di dati di altri fogli elettronici:

Tipo e larghezza dei dati. Il tipo e la larghezza dei dati di ciascuna variabile vengono determinati dalla larghezza della colonna e dal tipo di dati della prima cella di dati della colonna. I valori di altri tipi verranno convertiti in valori mancanti di sistema. Se la prima cella di dati della colonna è vuota, viene utilizzato il tipo di dati globale predefinito per il foglio elettronico (in genere numerico).

Celle vuote. Per le variabili numeriche, le celle vuote vengono convertite al valore mancante di sistema, indicato da un punto. Per le variabili stringa, uno spazio rappresenta un valore stringa valido. Le celle vuote, pertanto, vengono considerate come valori stringa validi.

Nomi delle variabili. Se i nomi delle variabili non vengono letti dal foglio elettronico, le lettere delle colonne (*A*, *B*, *C*, ...) verranno utilizzate come nomi delle variabili nei file di Excel e di Lotus. Per i file SYLK e i file di Excel salvati in formato di visualizzazione R1C1, per i nomi delle variabili (*C1*, *C2*, *C3*, ...) verrà utilizzato il numero di colonna preceduto dalla lettera *C*.

Lettura di file dBASE

Dal punto di vista logico, i file database sono molto simili ai file di dati in formato SPSS Statistics. Per i file di dBASE sono valide le seguenti regole generali:

- I nomi dei campi vengono convertiti in nomi validi di variabili.
- I due punti utilizzati nei nomi di campo di dBASE verranno tradotti in caratteri di sottolineatura.
- Verranno inclusi i record contrassegnati per l'eliminazione che non sono stati effettivamente rimossi. Per i casi contrassegnati per l'eliminazione, verrà creata una nuova variabile stringa, *D_R*, contenente un asterisco.

Lettura di file Stata

Per i file di Stata sono valide le seguenti regole generali:

- **Nomi delle variabili.** I nomi delle variabili di stratificazione vengono convertiti nei nomi delle variabili SPSS Statistics in formato che fa distinzione tra le maiuscole e le minuscole. I nomi delle variabili di stratificazione che sono identici, ad eccezione delle maiuscole/minuscole, vengono convertiti in nomi di variabili validi tramite l'aggiunta di un carattere di sottolineatura e una lettera sequenziale (*_A*, *_B*, *_C*, ..., *_Z*, *_AA*, *_AB*, ... e così via).
- **Etichette delle variabili.** Le etichette delle variabili di stratificazione vengono convertite in etichette delle variabili SPSS Statistics.
- **Etichette dei valori.** Le etichette dei valori di stratificazione vengono convertite in etichette dei valori SPSS Statistics, ad eccezione delle etichette dei valori di stratificazione assegnate a valori mancanti "estesi".
- **Valori mancanti.** I valori mancanti "estesi" di stratificazione vengono convertiti in valori mancanti di sistema.
- **Conversione di date.** I valori di stratificazione in formato di data vengono convertiti in valori SPSS Statistics in formato `DATE` (g-m-a). I valori di stratificazione in formato data "serie storica" (settimane, mesi, trimestri e così via) vengono convertiti nel formato numerico semplice (F), mantenendo il valore intero originale interno, ovvero il numero di settimane, mesi, trimestri e così via, a partire dal 1960.

Lettura di file di database

Si possono leggere dati in qualsiasi formato di database per cui si dispone del driver appropriato. Nella modalità di analisi locale, è necessario che nel computer locale siano installati i driver richiesti. Nella modalità di analisi distribuita (disponibile con SPSS Statistics Server), i driver devono essere installati sul server remoto.Per ulteriori informazioni, vedere Analisi distribuita in Capitolo 4 a pag. 69.

Per leggere i file di database

- Dai menu, scegliere:

  File
  Apri database
  Nuova query...

- Selezionare la sorgente dati.
- Se necessario (in base alla sorgente dati selezionata), specificare il file di database e/o di digitare un nome, una password e altre informazioni di login.
- Selezionare la tabella o le tabelle e i campi. Per le sorgenti dati OLE DB (disponibili solo nei sistemi operativi Windows), è possibile selezionare solo una tabella.
- Specificare le relazioni tra le tabelle.
- Oppure:
  - Specificare i criteri di selezione dei dati.
  - Aggiungere un prompt per l'input utente per creare una query di parametri.
  - Salvare la query prima di eseguirla.

Per modificare una query a database salvata

- Dai menu, scegliere:

  File
  Apri database
  Modifica query...

- Selezionare il file di query (*.*spq*) da modificare.
- Seguire le istruzioni per la creazione di una nuova query.

Per leggere file di database con query salvate

- Dai menu, scegliere:

  File
  Apri database
  Esegui query...

- Selezionare il file di query (*.*spq*) da eseguire.
- Se necessario (in base al file di database), immettere un nome e una password di login.
- Se la query ha un prompt incorporato, potrebbe risultare necessaria l'immissione di altre informazioni (ad esempio, il trimestre per cui si desidera ottenere le cifre relative alle vendite).

Selezione di una sorgente dati

Utilizzare questa prima schermata di Creazione guidata di query per selezionare il tipo di sorgente dati da leggere.

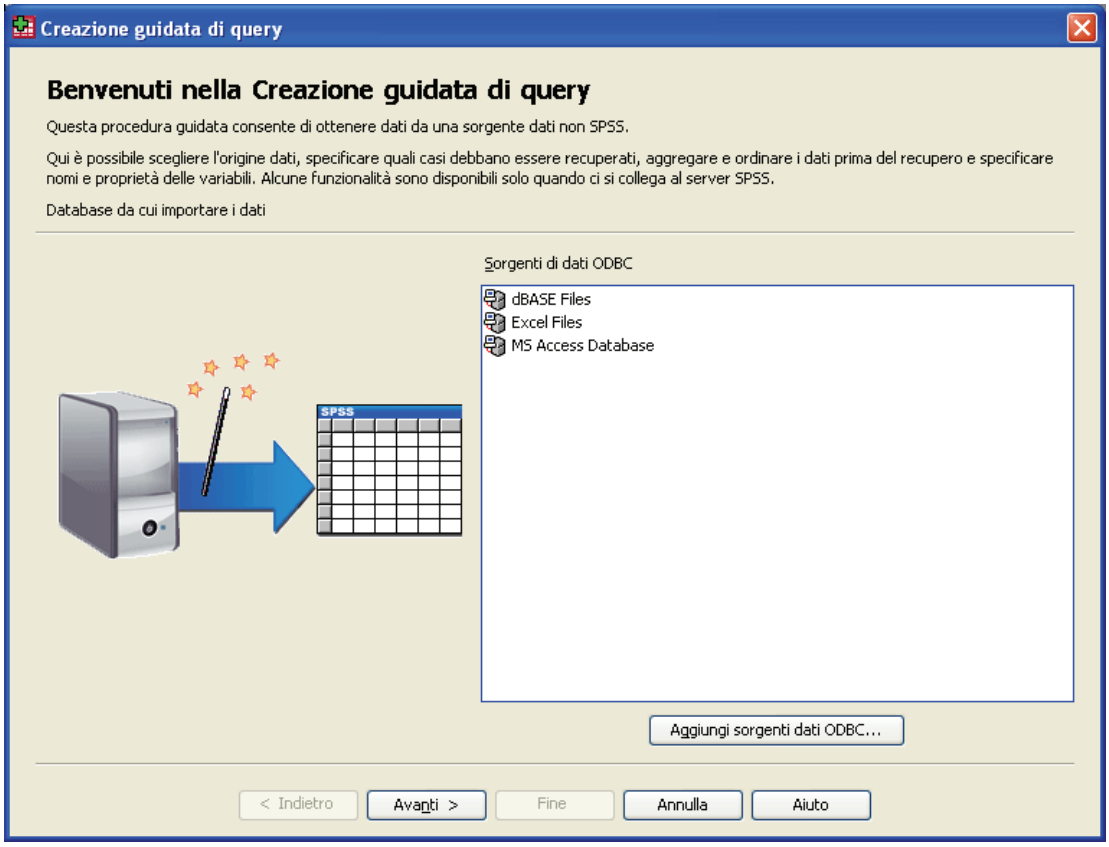
Sorgenti di dati ODBC

Se non è stata configurata alcuna sorgente dati ODBC o se si desidera aggiungerne una nuova, fare clic su Aggiungi sorgente dati ODBC.

- Nei sistemi operativi Linux, questo pulsante non è disponibile. Le sorgenti dati ODBC vengono specificate nel file *odbc.ini* e le variabili di ambiente *ODBCINI* devono essere impostate sulla posizione di tale file. Per ulteriori informazioni, vedere la documentazione relativa ai driver del database.
- Nella modalità di analisi distribuita (disponibile con SPSS Statistics Server) questo pulsante non è disponibile. Per aggiungere sorgenti dati nella modalità di analisi distribuita, rivolgersi all'amministratore di sistema.

Una sorgente dati ODBC include due informazioni fondamentali: il driver che verrà utilizzato per accedere ai dati e la posizione del database a cui si desidera accedere. Per specificare le sorgenti dati, è necessario che siano installati i driver appropriati Per la modalità di analisi locale, è possibile installare i driver per vari formati di database dal CD di installazione di SPSS Statistics.

Figura 3-1
Creazione guidata di query


Creazione guidata di query
Benvenuti nella Creazione guidata di query
Questa procedura guidata consente di ottenere dati da una sorgente dati non SPSS.
Qui è possibile scegliere l'origine dati, specificare quali casi debbano essere recuperati, aggregare e ordinare i dati prima del recupero e specificare nomi e proprietà delle variabili. Alcune funzionalità sono disponibili solo quando ci si collega al server SPSS.
Database da cui importare i dati
Sorgenti di dati ODBC
dBASE Files
Excel Files
MS Access Database
SPSS
Aggiungi sorgenti dati ODBC...
< Indietro
Avanti >
Fine
Annulla
Aiuto

Sorgenti dati OLE DB

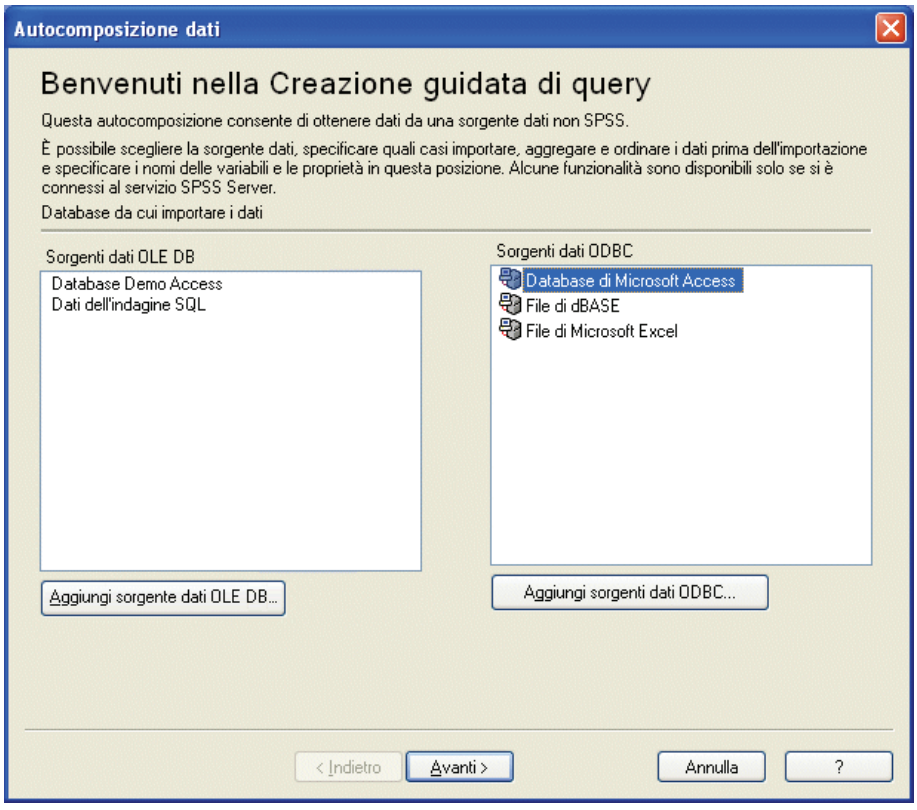
Per accedere alle sorgenti dati OLE DB (disponibili solo nei sistemi operativi Microsoft Windows), è necessario che siano installati i seguenti componenti:

- .NET framework
- Dimensions Data Model e OLE DB Access

Le versioni di tali componenti, compatibili con questa versione, possono essere installate dal CD di installazione e sono disponibili nel menu di esecuzione automatica.

- I join di tabelle non sono disponibili per le sorgenti dati OLE DB. È possibile leggere una sola tabella alla volta.
- È possibile aggiungere le sorgenti dati OLE DB solo in modalità di analisi locale. Per aggiungere delle sorgenti dati OLE DB in modalità di analisi distribuita in un server Windows, consultare l'amministratore di sistema.
- Nella modalità di analisi distribuita (disponibile con SPSS Statistics Server), le sorgenti dati OLE DB sono disponibili solo sui server Windows ed è necessario installare nel server .NET, Dimensions Data Model e OLE DB Access.

Figura 3-2
Creazione guidata di query con accesso alle sorgenti dati OLE DB


Autocomposizione dati
Benvenuti nella Creazione guidata di query
Questa autocomposizione consente di ottenere dati da una sorgente dati non SPSS.
È possibile scegliere la sorgente dati, specificare quali casi importare, aggregare e ordinare i dati prima dell'importazione e specificare i nomi delle variabili e le proprietà in questa posizione. Alcune funzionalità sono disponibili solo se si è connessi al servizio SPSS Server.
Database da cui importare i dati
Sorgenti dati OLE DB
Database Demo Access
Dati dell'indagine SQL
Aggiungi sorgente dati OLE DB...
Sorgenti dati ODBC
Database di Microsoft Access
File di dBASE
File di Microsoft Excel
Aggiungi sorgenti dati ODBC...
< Indietro
Avanti >
Annulla
?

Per aggiungere una sorgente dati OLE DB:

► Fare clic su Aggiungi sorgente dati OLE DB.

- In Proprietà di Data Link, fare clic sulla scheda Provider e selezionare il provider OLE DB.
- Fare clic su Avanti o sulla scheda Connessione.
- Selezionare il database immettendo la posizione della directory e il nome del database o facendo clic sul pulsante per selezionare un database. (Potrebbero essere necessari anche un nome utente e una password).
- Fare clic su OK dopo aver immesso le informazioni necessarie. (Per accertarsi che il database specificato sia disponibile, fare clic sul pulsante Prova connessione).
- Immettere un nome per le informazioni sulla connessione del database. (Tale nome verrà visualizzato nell'elenco delle sorgenti dati OLE DB disponibili).

Figura 3-3
Finestra di dialogo Salva con nome informazioni sulla connessione OLE DB

- Fare clic su OK.

L'operazione riconduce alla prima schermata di Creazione guidata di query, in cui è possibile selezionare il nome salvato dall'elenco di sorgenti dati OLE DB e proseguire con il passo successivo della procedura.

Eliminazione delle sorgenti dati OLE DB

Per eliminare i nomi delle sorgenti dati dall'elenco di sorgenti dati OLE DB, eliminare il file UDL che corrisponde al nome della sorgente dati contenuta in:

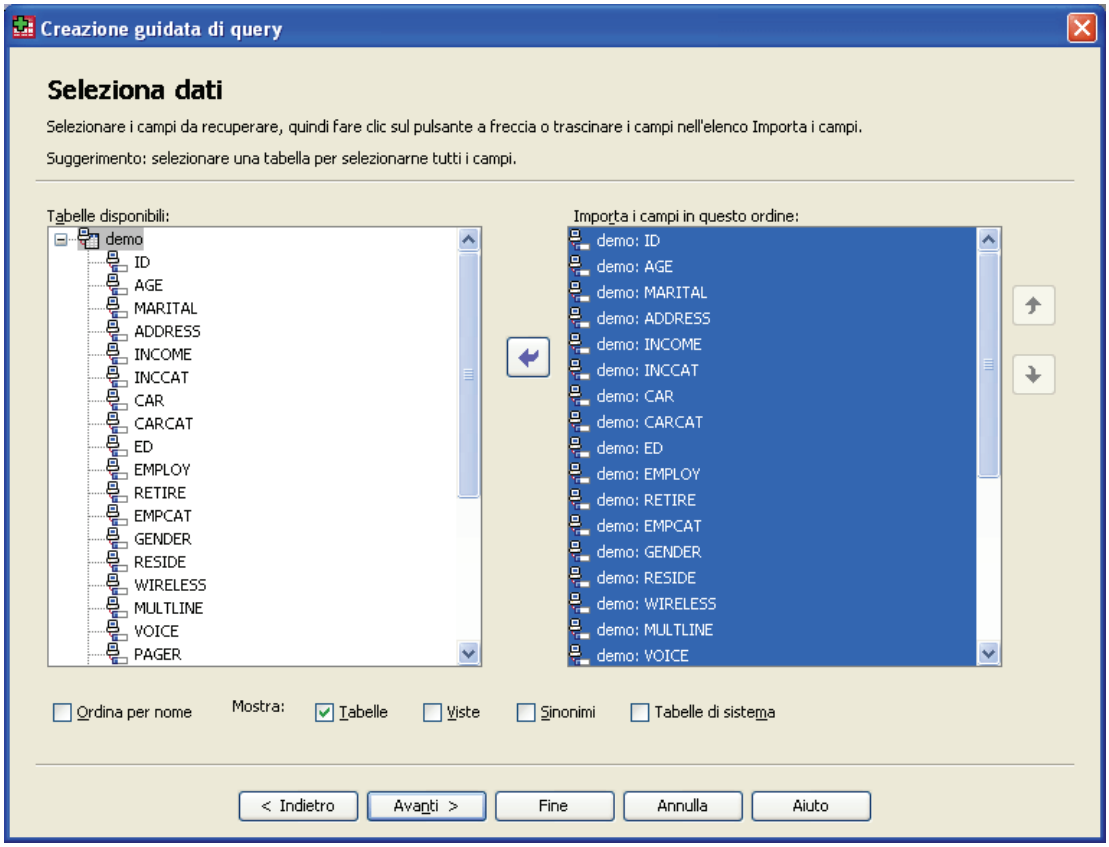
[unità]:\Documents and Settings\[accesso utente]\Impostazioni locali\Dati applicazioni\SPSS\UDL

Selezione dei campi di dati

Il passo Seleziona dei dati controlla le tabelle ed i campi che vengono letti. I campi del database (colonne) verranno letti come variabili.

Se in una tabella sono stati selezionati dei campi, tutti i campi saranno visibili nelle seguenti finestre della Creazione guidata di query, ma solo i campi selezionati in questo passo verranno importati come variabili. In questo modo è possibile creare join di tabelle e specificare i criteri utilizzando i campi che non vengono importati.

Figura 3-4
Creazione guidata di query, selezione dei dati



Visualizzazione dei nomi di campo. Per elencare i campi di una tabella, fare clic sul segno più (+) a sinistra del nome della tabella. Per nascondere i campi, fare clic sul segno meno (–), posto a sinistra del nome della tabella.

Per aggiungere un campo. Fare doppio clic su un campo qualsiasi nell'elenco Tabelle disponibili oppure trascinarlo nell'elenco Importa i campi in questo ordine. Per riordinare i campi, è possibile trascinarli nell'elenco di campi.

Per rimuovere un campo. Fare doppio clic su un campo nell'elenco Importa i campi in questo ordine oppure trascinarlo nell'elenco Tabelle disponibili.

Ordina per nome. Se la casella di controllo è selezionata, nella Creazione guidata di query verranno visualizzati i campi disponibili in ordine alfabetico.

Per impostazione predefinita, l'elenco delle tabelle disponibili visualizza solo le tabelle dei database standard. È possibile impostare il tipo di oggetti da visualizzare nell'elenco:

- **Tabelle.** Tabelle di database standard.
- **Viste.** Le viste sono "tabelle" virtuali o dinamiche definite da query. Possono includere di più tabelle e/o campi frutto di calcoli basati sui valori di altri campi.

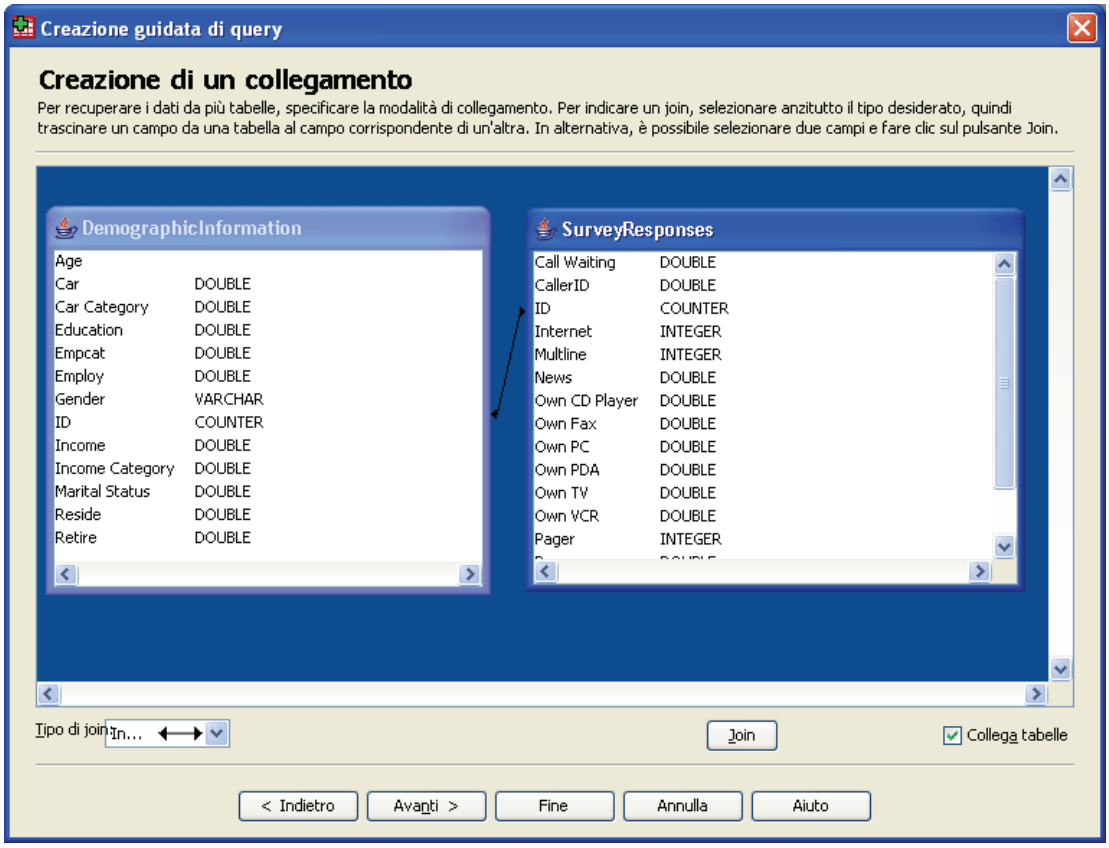
- **Sinonimi.** Un sinonimo è l'alias di una tabella o di una vista, di norma definito in una query.
- **Tabelle di sistema.** Le tabelle di sistema definiscono le proprietà del database. In alcuni casi, le tabelle di database standard sono classificate come tabelle di sistema e possono essere visualizzate solo selezionando questa opzione. Spesso solo agli amministratori possono accedere alle vere tabelle di sistema.

Nota: per le sorgenti dati OLE DB (disponibili solo nei sistemi operativi Windows), è possibile selezionare i campi solo da una singola tabella. I join di tabelle multiple non sono supportati per le sorgenti dati OLE DB.

Creazione di una relazione tra tabelle

Nel passo Creazione di un collegamento è possibile definire le relazioni tra le tabelle relative alle sorgenti dati ODBC. Se vengono selezionati campi da più tabelle, è necessario definire almeno un join.

Figura 3-5
Creazione guidata di query, definizione di relazioni



Creazione di relazioni. Per creare una relazione, trascinare un campo di una tabella nel campo a cui si desidera collegarlo. La Creazione guidata di query traccerà una **linea di collegamento** tra i due campi indicandone la relazione. I campi devono prevedere lo stesso tipo di dati.

Collega tabelle. Questa procedura consente di eseguire automaticamente un tentativo di collegamento delle tabelle mediante le chiavi principali ed esterne oppure confrontando i nomi del campo con il tipo di dati.

Tipo di join. Se il driver utilizzato supporta i join esterni, è possibile specificare join interni, join esterni sinistri e join esterni destri.

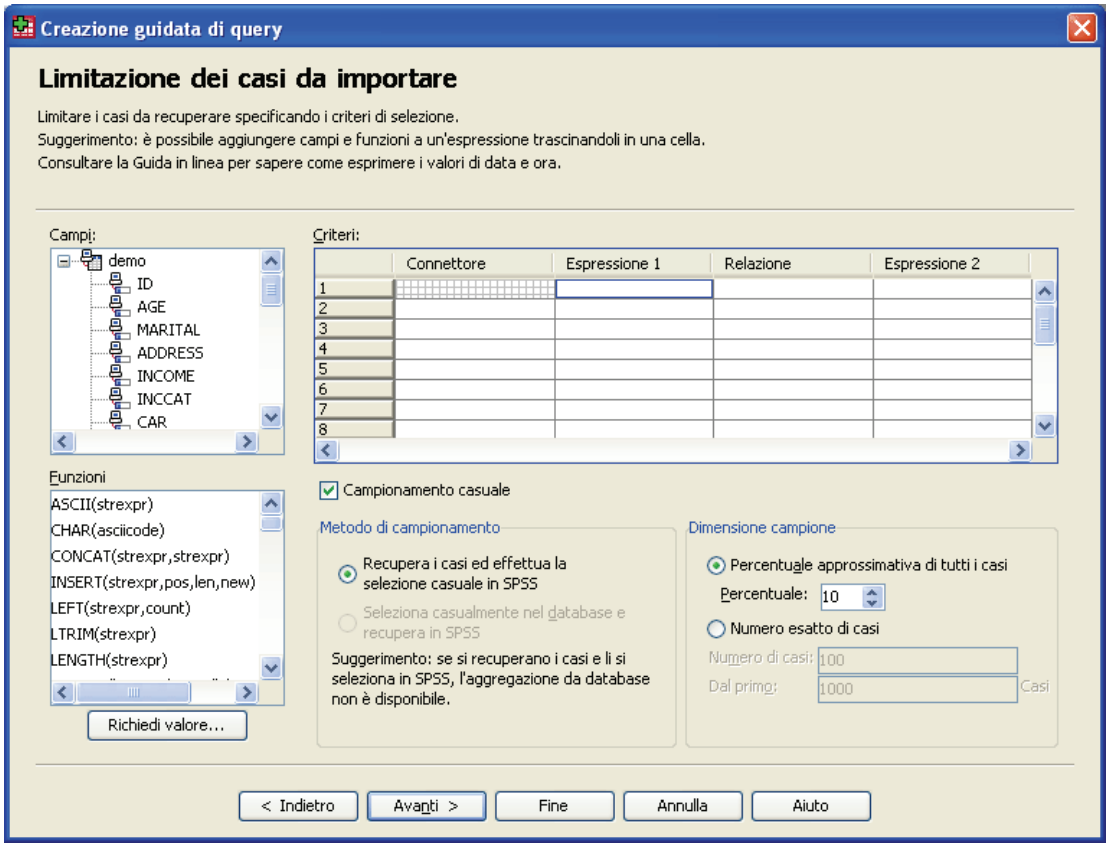
- **Join interni.** Un join interno include solo le righe in cui i campi correlati sono uguali. Nell'esempio, verranno incluse tutte le righe che riportano valori ID corrispondenti nelle due tabelle.
- **Join esterni.** Oltre alla corrispondenza di uno a uno con i join interni, è possibile utilizzare i join esterni per unire le tabelle con uno schema di corrispondenza di uno a molti. Per esempio, è possibile far corrispondere una tabella nella quale sono inclusi sono pochi record che rappresentano i valori dei dati e le etichette descrittive a essi associate, con dei valori di una tabella che contiene centinaia o migliaia di record che rappresentano i rispondenti del sondaggio. Un join esterno sinistro include tutti i record della tabella a sinistra e solo i record della tabella a destra in cui i campi correlati sono uguali. In un join esterno destro, il join importa tutti i i record della tabella a destra e solo i record della tabella a sinistra in cui i campi correlati sono uguali.

Limitazione dei casi da importare

Il passo Limitazione dei casi da importare consente di specificare i criteri per la selezione di sottoinsiemi di casi (righe). In genere, la limitazione dei casi consiste nel riempire la griglia dei criteri con dei criteri. I criteri consistono di due espressioni e di alcune relazioni reciproche. Le espressioni restituiscono un valore *vero*, *falso* o *mancante* per ciascun caso.

- Se il risultato è *vero*, il caso verrà selezionato.
- Se il risultato è *falso* o *mancante*, il caso non verrà selezionato.
- La maggior parte dei criteri utilizza uno o più dei sei operatori relazionali (<, >, <=, >=, = e <>).
- Le espressioni logiche possono includere nomi di campo, costanti, operatori aritmetici, funzioni numeriche e di altro tipo, nonché variabili logiche. È possibile utilizzare i campi che non si prevede di importare come variabili.

Figura 3-6
Creazione guidata di query, limitazione dei casi recuperati



Per creare criteri personalizzati sono necessarie almeno due espressioni e una relazione che colleghi le espressioni.

- Per creare un'espressione, scegliere uno dei seguenti metodi:
  - In una cella Espressione, immettere i nomi di campo, le costanti, gli operatori aritmetici e le funzioni numeriche e di altro tipo, nonché le variabili logiche.
  - Fare doppio clic sul campo nell'elenco Campi.
  - Trascinare il campo dall'elenco Campi in una cella Espressione.
  - Selezionare un campo dal menu a discesa in qualsiasi cella Espressione attiva.
- Per selezionare un operatore relazionale (ad esempio = o >), posizionare il cursore sulla cella Relazione e digitare l'operatore o selezionarlo dal menu a discesa.

Se l'istruzione SQL contiene proposizioni `WHERE` con espressioni per la selezione dei casi, è necessario specificare le date e le ore nelle espressioni in modo particolare (utilizzando anche le parentesi graffe indicate negli esempi):

- Le date in lettere devono essere specificate utilizzando il formato generico `{d 'aaaa-mm-gg'}`.
- Le ore in lettere devono essere specificate utilizzando il formato generico `{t 'hh:mm:ss'}`.

- Le date e le ore in lettere (contrassegni orari) devono essere specificate nel formato generico `{ts 'aaaa-mm-gg hh:mm:ss'}`.
- Il valore della data e/o dell'ora completo deve essere racchiuso da virgolette semplici. Gli anni devono essere specificati con il formato a quattro cifre e le date e le ore devono contenere due cifre per ciascuna parte del valore. Ad esempio la data 1° gennaio 2005, 1:05 AM verrebbe scritta nel seguente modo:

  ```
  {ts '2005-01-01 01:05:00'}
  ```

Funzioni. In SPSS sono disponibili diverse funzioni SQL predefinite di tipo aritmetico, logico, stringa, data e orario. È possibile trascinare una funzione dalla lista nell'espressione oppure inserire una funzione SQL valida. Per la lista delle funzioni SQL valide, consultare la documentazione del database. Per l'elenco delle funzioni standard disponibili, vedere:

http://msdn2.microsoft.com/en-us/library/ms711813.aspx

Usa campionamento casuale. Questa opzione consente di selezionare un campione casuale di casi dalla sorgente dati. Per le sorgenti dati di grandi dimensioni, è possibile limitare il numero dei casi a un campione piccolo e rappresentativo, riducendo in modo significativo i tempi di elaborazione. Il campionamento casuale nativo, se disponibile per la sorgente dati, risulta più veloce rispetto al campionamento casuale SPSS Statistics, perché il campionamento casuale SPSS Statistics deve leggere l'intera sorgente dati per estrarre un campione casuale.

- **Approssimativamente.** Genera un campione casuale che include approssimativamente la percentuale di casi specificata. Poiché per ciascun caso viene eseguito un processo indipendente di decisione pseudo-casuale, l'equivalenza tra la percentuale di casi selezionati e la percentuale specificata può essere solo approssimativa. Maggiore è il numero di casi inclusi nel data file e maggiore sarà l'approssimazione della percentuale di casi selezionati rispetto alla percentuale specificata.
- **Esattamente.** Seleziona un campione casuale del numero di casi specificato dal totale di casi specificato. Se il numero indicato è superiore al numero totale di casi nel file, il numero di casi estratti verrà ridotto proporzionalmente.

Nota: se si utilizza il campionamento casuale, l'aggregazione (disponibile in modalità distribuita in SPSS Statistics Server) non è disponibile.

Richiedi valore. È possibile inserire nella query la richiesta di creare una **query di parametri**. Quando un utente esegue la query, verrà richiesto di immettere le informazioni (in base a quanto è stato specificato). Ciò potrebbe risultare utile per visualizzare gli stessi dati in modi diversi. È possibile, ad esempio, eseguire la stessa query per visualizzare le cifre di vendita relative a diversi trimestri fiscali.

▶ Posizionare il cursore su una cella Espressione e fare clic su Richiedi valore per creare una richiesta.

Creazione di una query di parametri

Utilizzare il passo Richiedi valore per creare una finestra di dialogo in cui vengono richieste informazioni all'utente ogni volta che si esegue una query. Questa procedura risulta utile per eseguire query per dati provenienti dalla stessa sorgente dati utilizzando criteri diversi.

Figura 3-7
Richiedi valore

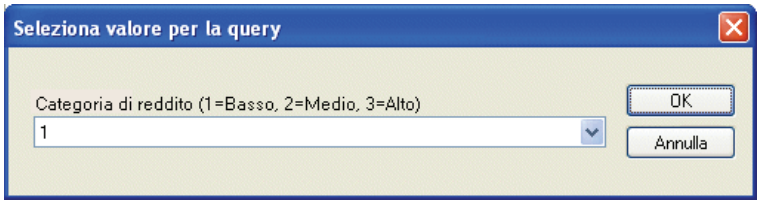
Per creare una richiesta di informazioni immettere una stringa di richiesta e un valore predefinito. Il testo della richiesta viene visualizzato ogni volta che un utente esegue la query. La stringa deve specificare il tipo di informazioni da immettere. Se l'utente non effettua la selezione da un elenco, la stringa deve fornire dei suggerimenti sulla formattazione dei dati. Esempio: Immettere un trimestre (t1, t2, t3, ...).

Scelta da elenco predefinito. Se la casella di controllo è selezionata, è possibile limitare i valori selezionabili dall'utente a quelli presenti in questo elenco. Accertarsi che i valori siano divisi da dei ritorni a capo.

Tipo dati. Selezionare il tipo di dati (numero, stringa o data).

Il risultato finale è simile al seguente:

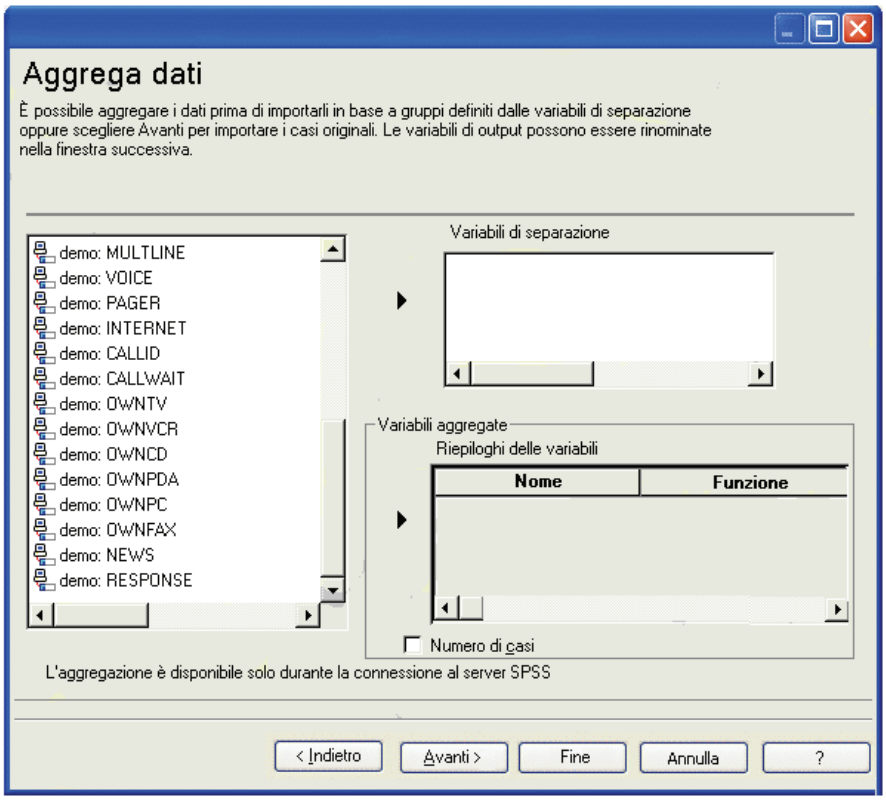
Figura 3-8
Prompt definito dall'utente



Aggregazione dei dati

Se è attiva la modalità di analisi distribuita in connessione con un server remoto (disponibile con SPSS Statistics Server), è possibile aggregare i dati prima di eseguirne la lettura in SPSS Statistics.

Figura 3-9
Creazione guidata di query, aggregazione dei dati



È inoltre possibile aggregare i dati dopo averne eseguito la lettura in SPSS Statistics; tuttavia, la pre-aggregazione può consentire di risparmiare tempo in caso di sorgenti dati di grandi dimensioni.

- Per creare dati aggregati, selezionare una o più variabili di separazione che definiscano le modalità di raggruppamento dei casi.
- Selezionare una o più variabili aggregate.
- Selezionare una funzione di aggregazione per ciascuna variabile aggregata.
- Se lo si desidera, è possibile creare una variabile che contenga il numero di casi di ciascun gruppo di interruzione.

  Nota: se si usa il campionamento casuale SPSS Statistics, l'aggregazione non è disponibile.

Per definire una variabile

Nomi di variabili ed etichette. Come etichetta di variabile verrà utilizzato il nome completo del campo del database (colonna). A meno che non venga modificato il nome della variabile, la Creazione guidata di query assegna i nomi di variabile a ciascuna colonna del database in uno dei due seguenti modi:

- Se il nome del campo del database forma un nome di variabile valido e univoco, questo verrà utilizzato come nome della variabile.
- Se il nome del campo del database non forma un nome di variabile valido e univoco, questo verrà creato automaticamente.

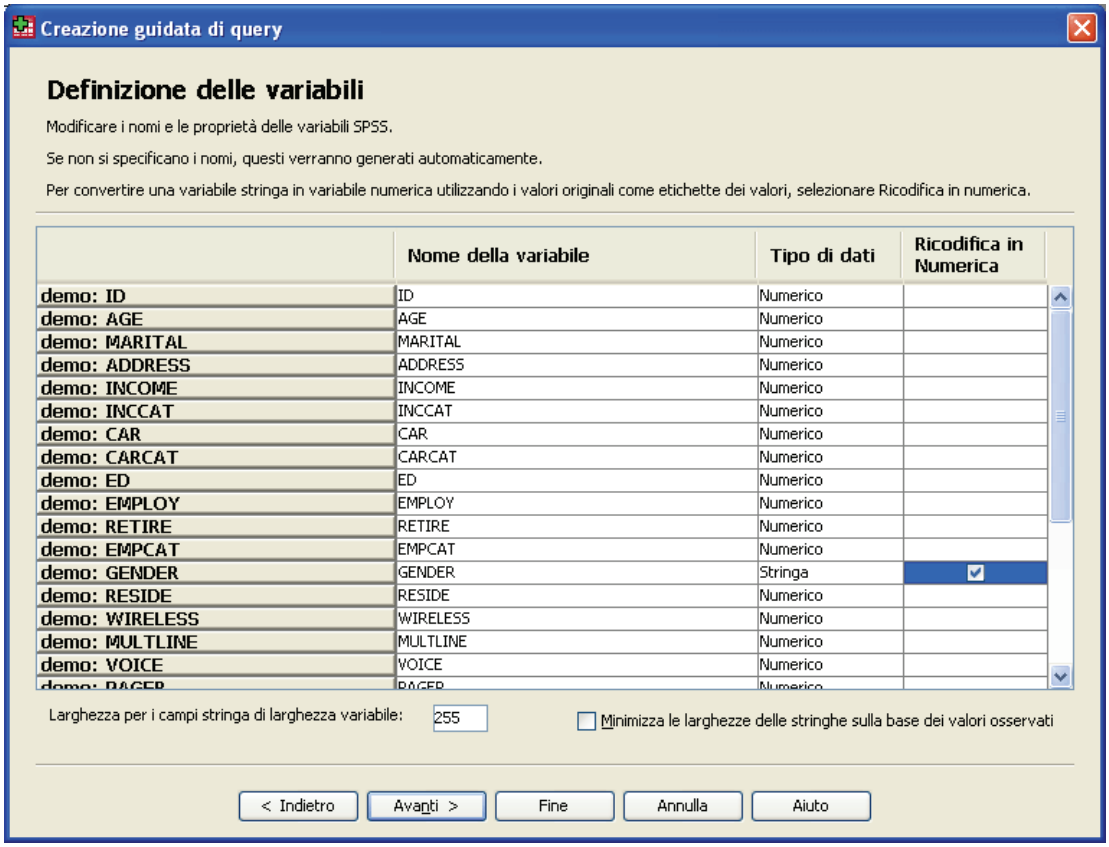
Fare clic su una cella per modificare il nome della variabile.

Conversione di stringhe in valori numerici. Selezionare la casella Ricodifica in numerica per la variabile stringa che si desidera convertire automaticamente in variabile numerica. I valori stringa verranno convertiti in valori interi consecutivi in base all'ordine alfabetico dei valori originali. I valori originali verranno mantenuti come etichette di valori per le nuove variabili.

Larghezza dei campi delle stringhe a larghezza variabile. L'opzione consente di controllare la larghezza dei valori delle stringhe a larghezza variabile. Per impostazione predefinita, la larghezza è 255 byte, sebbene vengano letti solo i primi 255 byte (generalmente i 255 caratteri nelle lingue a un byte). La larghezza massima è 32.767 byte. Sebbene sia preferibile non troncare i valori stringa, è consigliabile anche non specificare un valore alto se non è necessario perché questo renderà inefficiente l'elaborazione.

Minimizza le larghezze delle stringhe sulla base dei valori osservati. Impostare automaticamente la larghezza di ogni variabile di stringa sul valore più lungo osservato.

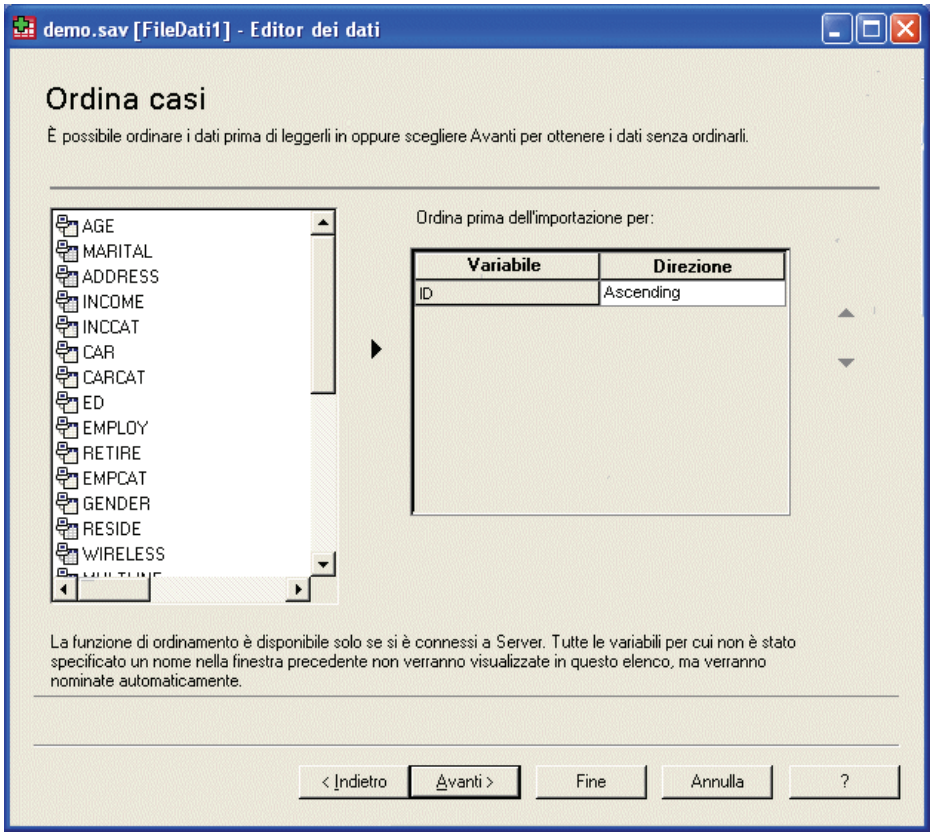
Figura 3-10
Creazione guidata di query, definizione di variabili


Creazione guidata di query
Definizione delle variabili
Modificare i nomi e le proprietà delle variabili SPSS.
Se non si specificano i nomi, questi verranno generati automaticamente.
Per convertire una variabile stringa in variabile numerica utilizzando i valori originali come etichette dei valori, selezionare Ricodifica in numerica.
Nome della variabile
Tipo di dati
Ricodifica in Numerica
demo: ID ID Numerico
demo: AGE AGE Numerico
demo: MARITAL MARITAL Numerico
demo: ADDRESS ADDRESS Numerico
demo: INCOME INCOME Numerico
demo: INCCAT INCCAT Numerico
demo: CAR CAR Numerico
demo: CARCAT CARCAT Numerico
demo: ED ED Numerico
demo: EMPLOY EMPLOY Numerico
demo: RETIRE RETIRE Numerico
demo: EMPCAT EMPCAT Numerico
demo: GENDER GENDER Stringa
demo: RESIDE RESIDE Numerico
demo: WIRELESS WIRELESS Numerico
demo: MULTLINE MULTLINE Numerico
demo: VOICE VOICE Numerico
Larghezza per i campi stringa di larghezza variabile: 255
Minimizza le larghezze delle stringhe sulla base dei valori osservati
< Indietro
Avanti >
Fine
Annulla
Aiuto

Ordinamento dei casi

Se è attiva la modalità di analisi distribuita in connessione con un server remoto (disponibile con SPSS Statistics Server), è possibile ordinare i dati prima di eseguirne la lettura in SPSS Statistics.

Figura 3-11
Creazione guidata di query, ordinamento di casi



È inoltre possibile ordinare i dati dopo averne eseguito la lettura in SPSS Statistics; tuttavia, il pre-ordinamento può consentire di risparmiare tempo in caso di sorgenti dati di grandi dimensioni.

Risultati

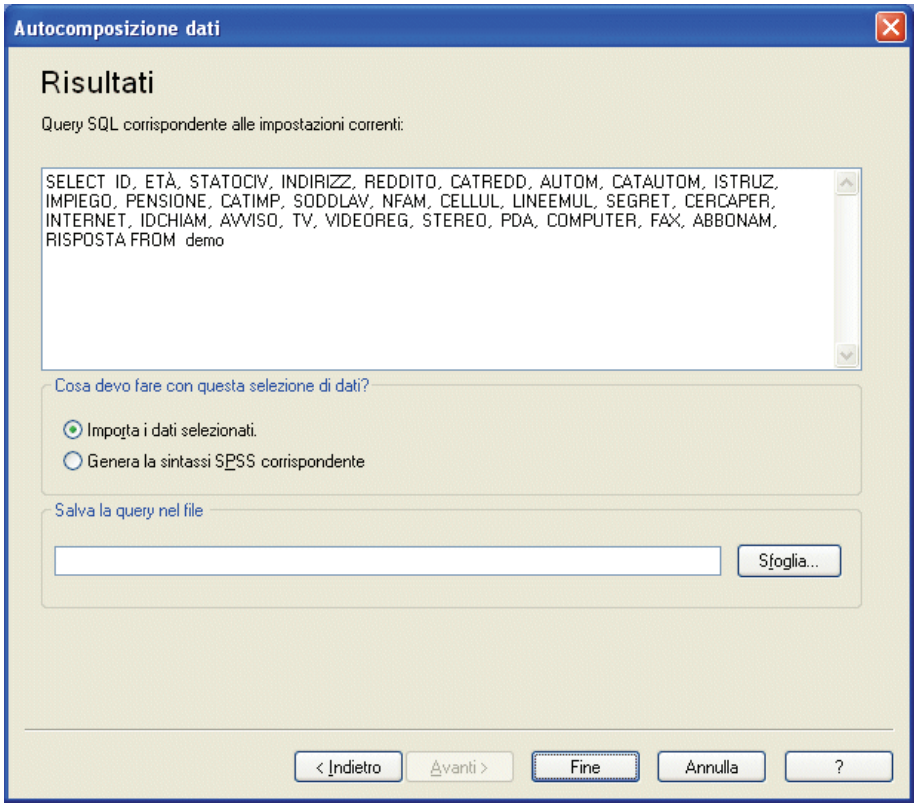
Nel passo Risultati viene visualizzata l'istruzione SQL Select della query.

- È possibile modificare l'istruzione SQL Select prima di eseguire la query, ma se si fa clic sul pulsante Indietro per apportare modifiche nei passi precedenti, le modifiche all'istruzione Select verranno perse.
- Per salvare la query e renderla disponibile per utilizzi futuri, utilizzare la sezione Salva la query nel file.
- Per incollare la sintassi completa di `GET DATA` in una finestra di sintassi, selezionare Genera la sintassi SPSS corrispondente. Se si copia e si incolla l'istruzione Select dalla finestra Risultati, la sintassi dei comandi richiesta non verrà incollata.

Nota: la sintassi incollata contiene uno spazio bianco prima della chiusura delle virgolette su ciascuna riga dell'SQL generato dalla procedura guidata. Questi spazi bianchi non sono superflui. Al momento dell'elaborazione del comando, tutte le righe dell'istruzione SQL vengono unite

letteralmente. Se non ci fosse lo spazio, non ci sarebbe alcuna divisione tra l'ultimo carattere della prima riga e il primo carattere di quella successiva.

Figura 3-12
Creazione guidata di query, riquadro risultati


Autocomposizione dati
Risultati
Query SQL corrispondente alle impostazioni correnti:
SELECT ID, ETÀ, STATOCIV, INDIRIZZ, REDDITO, CATREDD, AUTOM, CATAUTOM, ISTRUZ, IMPIEGO, PENSIONE, CATIMP, SODDLAV, NFAM, CELLUL, LINEEMUL, SEGRET, CERCAPER, INTERNET, IDCHIAM, AVVISO, TV, VIDEOREG, STEREO, PDA, COMPUTER, FAX, ABBONAM, RISPOSTA FROM demo
Cosa devo fare con questa selezione di dati?
Importa i dati selezionati.
Genera la sintassi SPSS corrispondente
Salva la query nel file
Sfoglia...
< Indietro
Avanti >
Fine
Annulla
?

Importazione guidata di testo

Importazione guidata di testo consente di leggere i file di testo formattati in vari modi:

- File tab delimitati
- File delimitati da spazi
- File delimitati da virgole
- File in formato di campo fisso

Per i file delimitati è inoltre possibile specificare altri caratteri come delimitatori da inserire tra i valori ed è possibile specificare più delimitatori.

Per leggere i file di dati di testo

► Dai menu, scegliere:

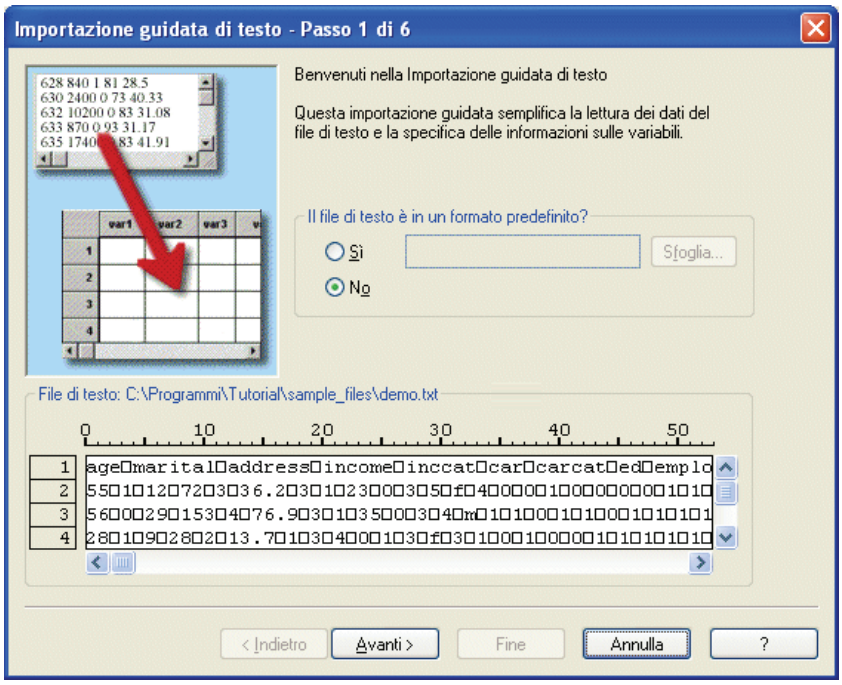
File
  Leggi dati testo...

► Selezionare il file di testo nella finestra di dialogo Apri.

▶ Eseguire i passi di Importazione guidata di testo per definire il metodo di lettura dei file.

Importazione guidata di testo: Passo 1

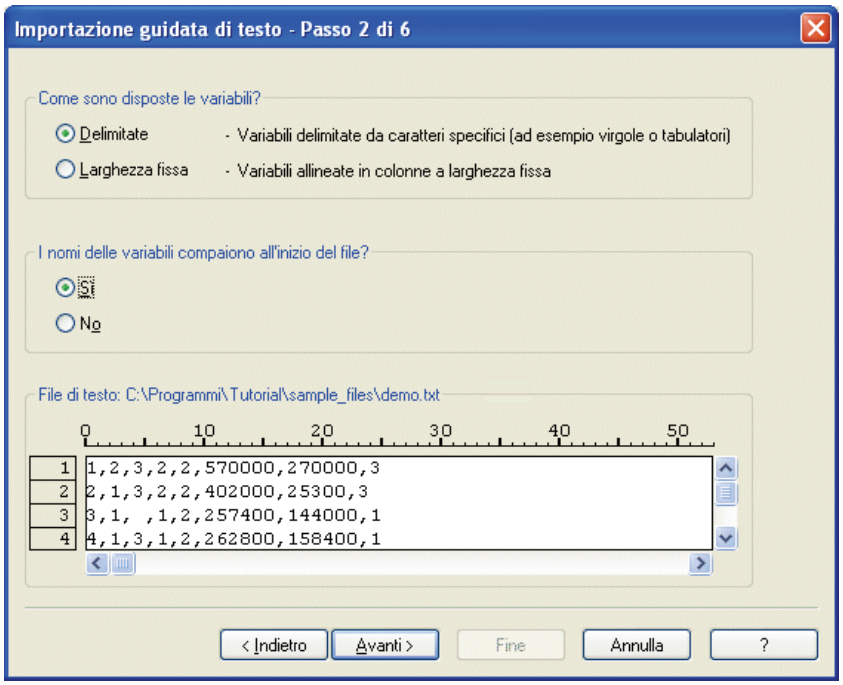
Figura 3-13
Importazione guidata di testo: Passo 1


Importazione guidata di testo - Passo 1 di 6
Benvenuti nella Importazione guidata di testo
Questa importazione guidata semplifica la lettura dei dati del file di testo e la specifica delle informazioni sulle variabili.
Il file di testo è in un formato predefinito?
Sì
No
Sfoglia...
File di testo: C:\Programmi\Tutorial\sample_files\demo.txt
< Indietro
Avanti >
Fine
Annulla
?

Il file di testo viene visualizzato in una finestra di anteprima. È possibile applicare un formato predefinito (salvato precedentemente con Importazione guidata di testo) o eseguire i passi della procedura guidata per definire il metodo di lettura dei dati.

Importazione guidata di testo: Passo 2

Figura 3-14
Importazione guidata di testo: Passo 2


Importazione guidata di testo - Passo 2 di 6
Come sono disposte le variabili?
Delimitate - Variabili delimitate da caratteri specifici (ad esempio virgole o tabulatori)
Larghezza fissa - Variabili allineate in colonne a larghezza fissa
I nomi delle variabili compaiono all'inizio del file?
Sì
No
File di testo: C:\Programmi\Tutorial\sample_files\demo.txt
0 10 20 30 40 50
1 1,2,3,2,2,570000,270000,3
2 2,1,3,2,2,402000,25300,3
3 3,1, ,1,2,257400,144000,1
4 4,1,3,1,2,262800,158400,1
< Indietro
Avanti >
Fine
Annulla
?

Nel passo 2 vengono fornite informazioni sulle variabili. Le variabili sono simili ai campi di un database. Ad esempio, ogni voce di un questionario è una variabile.

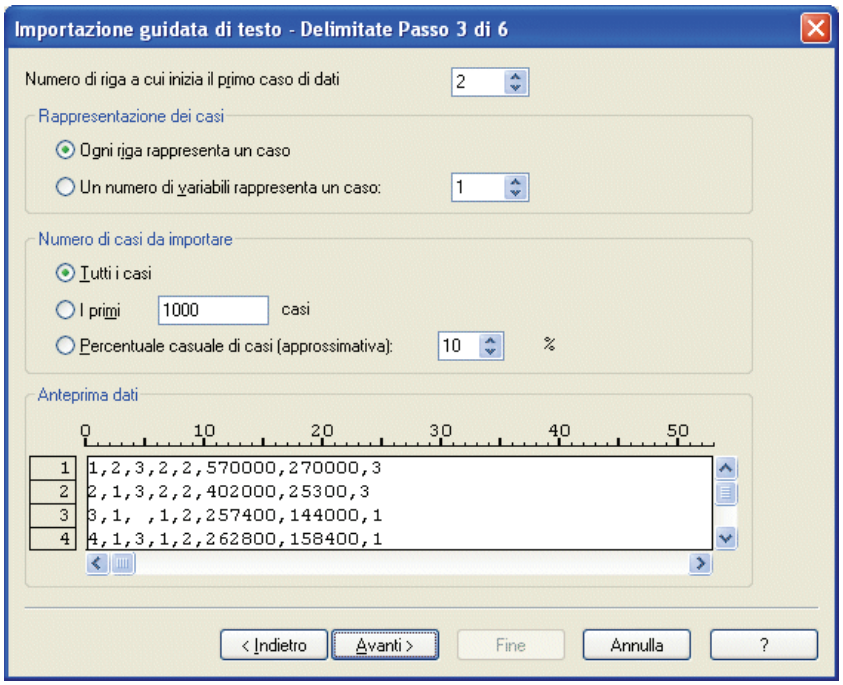
Come sono disposte le variabili? Per leggere correttamente i dati, Importazione guidata di testo deve poter determinare dove finiscono i valori dei dati di una variabile e iniziano i valori di quella successiva. La disposizione delle variabili determina il metodo utilizzato per separare una variabile dalla successiva.

- **Delimitate.** Per separare le variabili possono essere usati spazi, virgole, tabulazioni o altri caratteri. Le variabili di ciascun caso vengono registrate nello stesso ordine, ma non sempre nelle stesse posizioni di colonna.
- **Larghezza fissa.** Ciascuna variabile viene registrata nella stessa posizione della colonna e nello stesso record (riga) per ciascun caso del file di dati. Non è richiesto nessun delimitatore tra le variabili. In molti file di dati di testo generati automaticamente potrebbe sembrare che i valori dei dati siano disposti senza alcun simbolo di divisione. La posizione della colonna determina la variabile che verrà letta.

I nomi delle variabili compaiono all'inizio del file? Se la prima riga del file di dati contiene etichette descrittive delle variabili, è possibile utilizzare queste etichette come nomi di variabili. I valori non conformi alle regole di denominazione delle variabili vengono convertiti in nomi di variabile validi.

Importazione guidata di testo: Passo 3 (File delimitati)

Figura 3-15
Importazione guidata di testo: Passo 3 (per file delimitati)



Nel passo 3 vengono fornite informazioni sui casi. I casi sono simili ai record di un database. Ad esempio, ogni persona che risponde a un questionario rappresenta un caso.

Numero di riga a cui inizia il primo caso di dati. Indica la prima riga del file di dati che contiene valori. Se la prima riga o le prime righe del file di dati contengono etichette descrittive o testo di altro tipo che non rappresenta un valore, questa *non*sarà la riga 1.

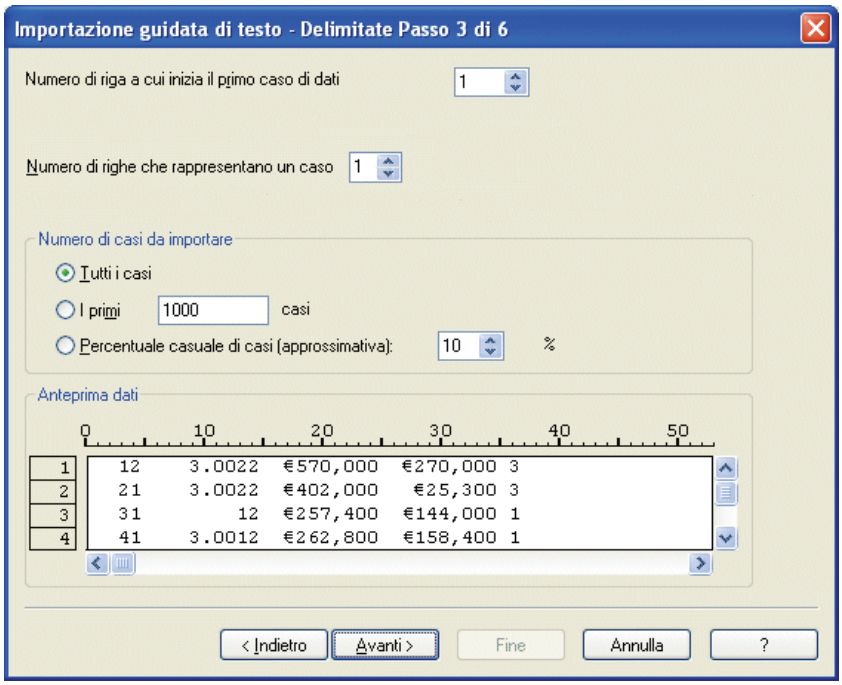
Rappresentazione dei casi. Controlla il modo in cui Importazione guidata di testo determina la fine di ogni caso e l'inizio di quello successivo.

- **Ogni riga rappresenta un caso.** Ogni riga contiene un solo caso. È abbastanza comune che ogni caso sia inserito su una sola riga, anche se questa potrebbe essere particolarmente lunga per i file di dati con un elevato numero di variabili. Se non tutte le righe contengono lo stesso numero di valori, il numero di variabili per ciascun caso è determinato dalla riga con il maggior numero di valori. Ai casi con meno valori sono assegnati valori mancanti per le variabili aggiuntive.
- **Un numero di variabili rappresenta un caso.** Il numero di variabili specificato per ogni caso indica a Importazione guidata di testo dove interrompere la lettura di un caso e iniziare quella del caso successivo. È possibile che più casi siano contenuti nella stessa riga e i casi possono anche iniziare a metà di una riga e proseguire nella riga successiva. Importazione guidata di testo determina la fine di ogni caso in base al numero di valori letti, indipendentemente dal numero di righe. Ogni caso deve contenere i valori dei dati o i valori mancanti indicati dai delimitatori per tutte le variabili, altrimenti i file di dati verranno letti in modo non corretto.

Numero di casi da importare. È possibile importare tutti i casi del file di dati, solo i primi *n* casi (*n* è un numero specificato dall'utente) o un campione casuale determinato in base a una percentuale specificata. Poiché viene eseguito un processo indipendente di decisione pseudo-casuale per ciascun caso, l'equivalenza tra la percentuale di casi selezionati e la percentuale specificata può essere solo approssimativa. Maggiore è il numero di casi inclusi nel data file e maggiore sarà l'approssimazione della percentuale di casi selezionati rispetto alla percentuale specificata.

Importazione guidata di testo: Passo 3 (File a larghezza fissa)

Figura 3-16
Importazione guidata di testo: Passo 3 (per file a larghezza fissa)



Nel passo 3 vengono fornite informazioni sui casi. I casi sono simili ai record di un database. Ad esempio, ogni persona che risponde a un questionario rappresenta un caso.

Numero di riga a cui inizia il primo caso di dati. Indica la prima riga del file di dati che contiene valori. Se la prima riga o le prime righe del file di dati contengono etichette descrittive o testo di altro tipo che non rappresenta un valore, questa *non*sarà la riga 1.

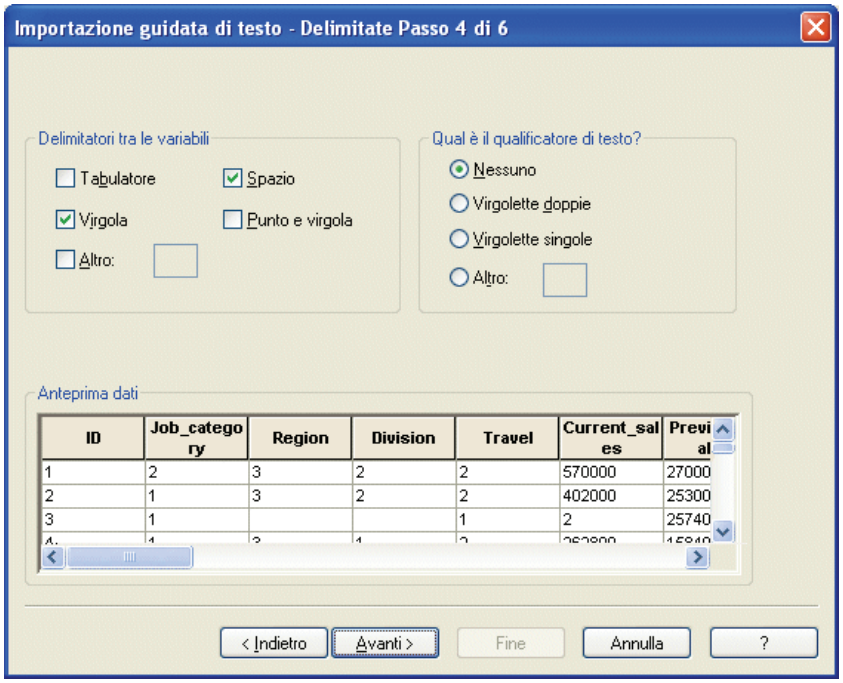
Numero di righe che rappresentano un caso. Controlla il modo in cui Importazione guidata di testo determina la fine di ogni caso e l'inizio di quello successivo. Ogni variabile è definita dal relativo numero di riga all'interno del caso e dalla posizione di colonna. È necessario specificare il numero di righe per ogni caso per poter leggere correttamente i dati.

Numero di casi da importare. È possibile importare tutti i casi del file di dati, solo i primi *n* casi (*n* è un numero specificato dall'utente) o un campione casuale determinato in base a una percentuale specificata. Poiché viene eseguito un processo indipendente di decisione pseudo-casuale per ciascun caso, l'equivalenza tra la percentuale di casi selezionati e la percentuale specificata può

essere solo approssimativa. Maggiore è il numero di casi inclusi nel data file e maggiore sarà l'approssimazione della percentuale di casi selezionati rispetto alla percentuale specificata.

Importazione guidata di testo: Passo 4 (File delimitati)

Figura 3-17
Importazione guidata di testo: Passo 4 (per file delimitati)



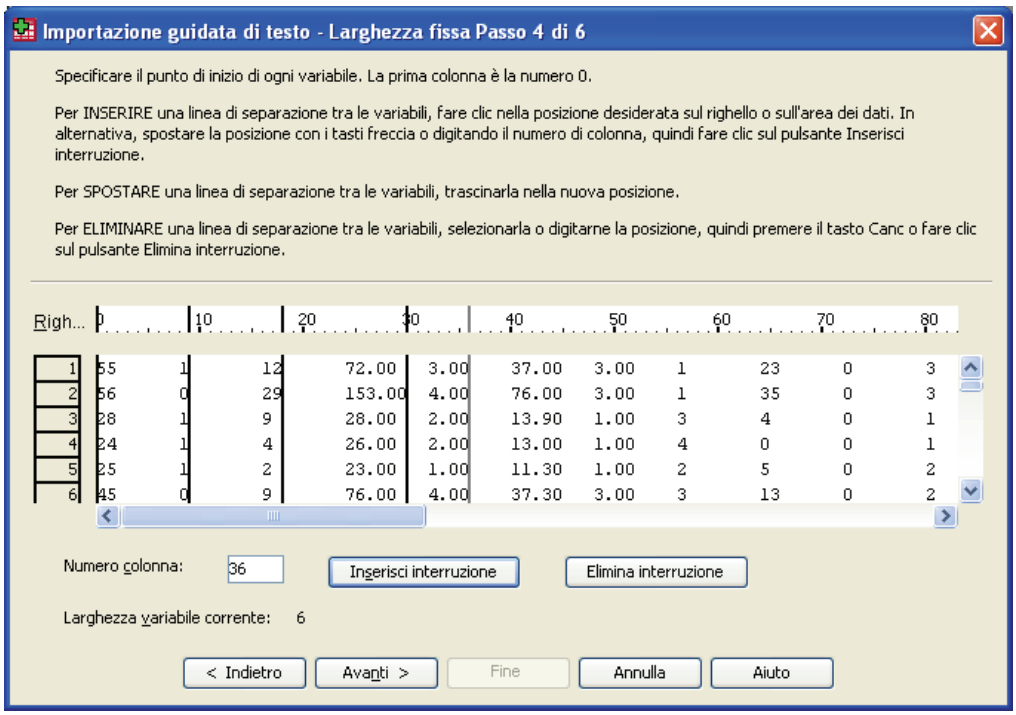
Questo passo illustra la soluzione più opportuna usata da Importazione guidata di testo per leggere un file di dati e consente di modificare il metodo impiegato per leggere le variabili dal file di dati.

Delimitatori tra le variabili. Indica i caratteri o i simboli che separano i valori. È possibile scegliere qualsiasi combinazione di spazi, virgole, punti e virgola, tabulazioni o caratteri di altro tipo. Se sono presenti più delimitatori consecutivi, questi sono considerati valori mancanti.

Qual è il qualificatore di testo? Caratteri usati per includere valori contenenti caratteri delimitatori. Ad esempio, se il delimitatore è costituito da una virgola, i valori che contengono virgole verranno letti in modo inesatto, a meno che non esista un qualificatore di testo contenente il valore e che impedisca alle virgole presenti nel valore di essere interpretate come delimitatori. I file dati in formato CSV esportati da Excel utilizzano le virgolette doppie (") come qualificatore di testo. Il qualificatore di testo viene visualizzato sia all'inizio sia alla fine del valore, racchiudendo in tal modo l'intero valore.

Importazione guidata di testo: Passo 4 (File a larghezza fissa)

Figura 3-18
Importazione guidata di testo: Passo 4 (per file a larghezza fissa)


Importazione guidata di testo - Larghezza fissa Passo 4 di 6
Specificare il punto di inizio di ogni variabile. La prima colonna è la numero 0.
Per INSERIRE una linea di separazione tra le variabili, fare clic nella posizione desiderata sul righello o sull'area dei dati. In alternativa, spostare la posizione con i tasti freccia o digitando il numero di colonna, quindi fare clic sul pulsante Inserisci interruzione.
Per SPOSTARE una linea di separazione tra le variabili, trascinarla nella nuova posizione.
Per ELIMINARE una linea di separazione tra le variabili, selezionarla o digitarne la posizione, quindi premere il tasto Canc o fare clic sul pulsante Elimina interruzione.
Righ...
Numero colonna:
36
Inserisci interruzione
Elimina interruzione
Larghezza variabile corrente:
6
< Indietro
Avanti >
Fine
Annulla
Aiuto

Questo passo illustra la soluzione più opportuna usata da Importazione guidata di testo per leggere un file di dati e consente di modificare il metodo impiegato per leggere le variabili dal file di dati. Le linee verticali nella finestra di anteprima indicano le posizioni in cui la procedura Importazione guidata di testo ipotizza possa iniziare una variabile.

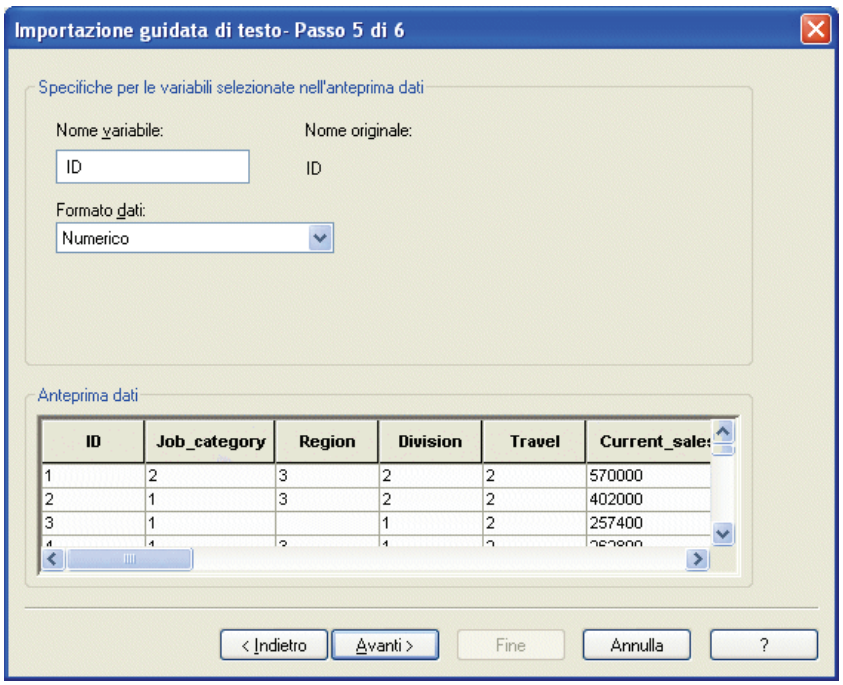
Inserire, spostare ed eliminare le linee di interruzione per separare le variabili in base alle necessità. Se per ogni caso vengono usate più righe, i dati verranno visualizzati come una riga per ogni caso, con le righe successive aggiunte alla fine della riga.

Note:

Per i file di dati generati dal computer che producono un flusso continuo di valori non intercalati da spazi o altri caratteri di separazione, potrebbe essere difficile determinare l'inizio di ogni variabile. Questi file in genere si basano su file di definizione dei dati o su qualche altra descrizione scritta che indica la posizione di riga e colonna per ogni variabile.

Importazione guidata di testo: Passo 5

Figura 3-19
Importazione guidata di testo: Passo 5



Questi passi controllano il nome della variabile e il formato dei dati usati da Importazione guidata di testo per leggere le variabili e quali variabili dovranno essere incluse nel file di dati finale.

Nome variabile. È possibile sovrascrivere i nomi predefiniti delle variabili con nomi personalizzati. Se i nomi delle variabili vengono letti dal file di dati, Importazione guidata di testo modificherà automaticamente i nomi delle variabili non conformi alle regole di denominazione delle variabili. Selezionare una variabile nella finestra di anteprima e specificarne il nome.

Formato dati. Selezionare una variabile dalla finestra di anteprima e quindi un formato dall'elenco a discesa. Fare clic tenendo premuto Maiusc per selezionare più variabili contigue oppure fare clic tenendo premuto Ctrl per selezionare più variabili non contigue.

Opzioni di formattazione di Importazione guidata di testo

Le opzioni di formattazione per la lettura delle variabili con Importazione guidata di testo includono:

Non importare. Omette le variabili selezionate dal file di dati importato.

Numerica. I valori validi possono essere numeri, un segno più (+) o meno (-) iniziale e un separatore decimale.

Stringa. I valori validi sono rappresentati da qualsiasi carattere incluso nella tastiera e dagli spazi interposti. Per i file delimitati, è possibile specificare il numero di caratteri per valore, fino a un massimo di 32,767. Per impostazione predefinita, Importazione guidata di testo imposta il numero di caratteri in base al valore della stringa più lunga tra le variabili selezionate. Per i

file a lunghezza fissa, il numero di caratteri nei valori stringa è determinato dalla posizione delle linee di interruzione delle variabili come definite nel passo 4.

Data/ora. I valori ammessi includono le date nel formato generale *gg-mm-aaaa*, *mm/gg/aaaa*, *gg.mm.aaaa*, *aaaa/mm/gg*, *hh:mm:ss*, nonché in molti altri formati. I mesi possono essere rappresentati in cifre, numeri romani, abbreviazioni di tre lettere o espressi con il nome completo. Selezionare un formato di data dall'elenco.

Dollaro. I valori validi sono numeri con un segno di dollaro iniziale (facoltativo) e la virgola (facoltativa) come separatore delle migliaia.

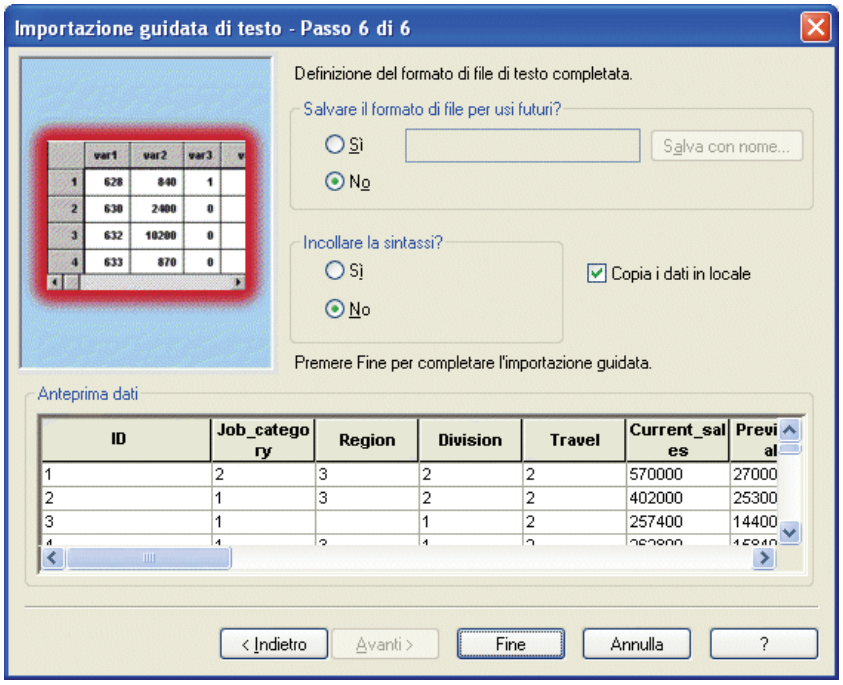
Virgola. Sono ammessi i numeri che usano un punto come separatore dei decimali e la virgola come separatore delle migliaia.

Punto. Sono ammessi i numeri che usano una virgola come separatore dei decimali e il punto come separatore delle migliaia.

Nota: I valori che contengono caratteri incompatibili con il formato selezionato saranno considerati valori mancanti. I valori che contengono uno qualsiasi dei separatori specificati verranno considerati valori multipli.

Importazione guidata di testo: Passo 6

Figura 3-20
Importazione guidata di testo: Passo 6



È il passo conclusivo della procedura. È possibile salvare le impostazioni in un file da utilizzare per l'importazione di file di dati di testo simili. È inoltre possibile incollare nella finestra appropriata la sintassi generata da Importazione guidata di testo. Infine è possibile personalizzare e/o salvare la sintassi in modo da poterla utilizzare in altre sessioni o durante le sessioni Utilità SPSS Production.

Copia i dati in locale. Una copia cache dei dati è una copia dell'intero file di dati, memorizzata in uno spazio su disco temporaneo. La copia dei dati può migliorare le prestazioni.

Lettura di dati Dimensions

Nei sistemi operativi Microsoft Windows è possibile leggere dati da prodotti Dimensions, compresi Quanvert, Quancept e mrInterview. (*Nota*: questa funzione è disponibile solo con SPSS Statistics installato sui sistemi operativi Microsoft Windows.

Per la lettura di sorgenti dati Dimensions, è necessario che siano installati i seguenti componenti:

- .NET framework
- Dimensions Data Model e OLE DB Access

Le versioni di tali componenti, compatibili con questa versione, possono essere installate dal CD di installazione e sono disponibili nel menu di esecuzione automatica. È possibile leggere le sorgenti dati Dimensions solo in modalità di analisi locale. Questa funzionalità non risulta disponibile in modalità di analisi distribuita quando si utilizza SPSS Statistics Server.

Per leggere i dati da una sorgente di dati Dimensions:

▶ Dai menu di qualsiasi finestra SPSS Statistics aperta, scegliere:
File
Apri dati Dimensions

▶ Nella scheda Connessione della finestra di dialogo Proprietà di Data Link, specificare i file di metadati, il tipo di dati del caso e il file di dati del caso.

▶ Fare clic su OK.

▶ Nella finestra di dialogo Importazione dati Dimensions selezionare le variabili che si desidera includere e tutti i criteri di selezione del caso.

▶ Fare clic su OK per leggere i dati.

Scheda Connessione della finestra di dialogo Proprietà di Data Link

Per leggere una sorgente di dati Dimensions, è necessario specificare:

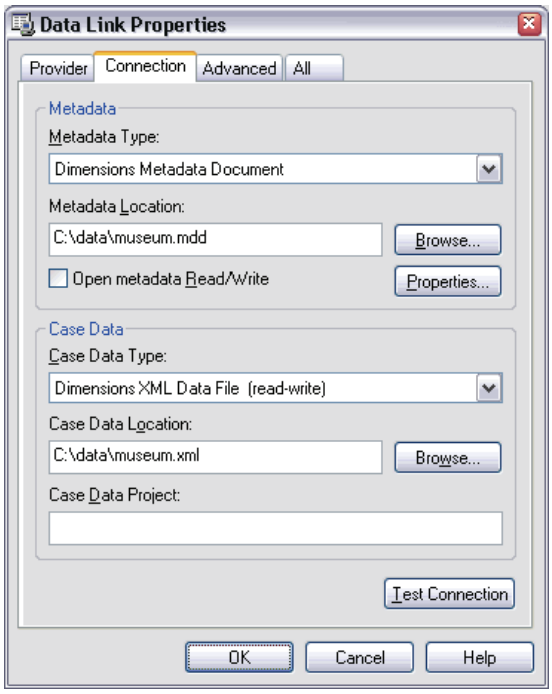
Posizione dei metadati. File documento dei metadati (*.mdd*) contenente informazioni sulla definizione del questionario.

Tipo di dati del caso. Formato del file di dati del caso. I formati disponibili includono:

- File di dati Quancept (DRS). Dati del caso in un file Quancept *.drs*, *.drz* o *.dru*.
- Database Quanvert. Dati del caso in un database Quanvert.
- Database Dimensions (MS SQL Server). Dati del caso in un database relazionale in SQL Server. Questa opzione può essere utilizzata per leggere i dati raccolti con mrInterview.
- File di dati XML Dimensions Dati del caso in un file XML.

Posizione dei dati del caso. File che contiene i dati del caso. Il formato di questo file deve essere coerente con il tipo di dati del caso selezionato.

Figura 3-21
Proprietà di Data Link: scheda Connessione



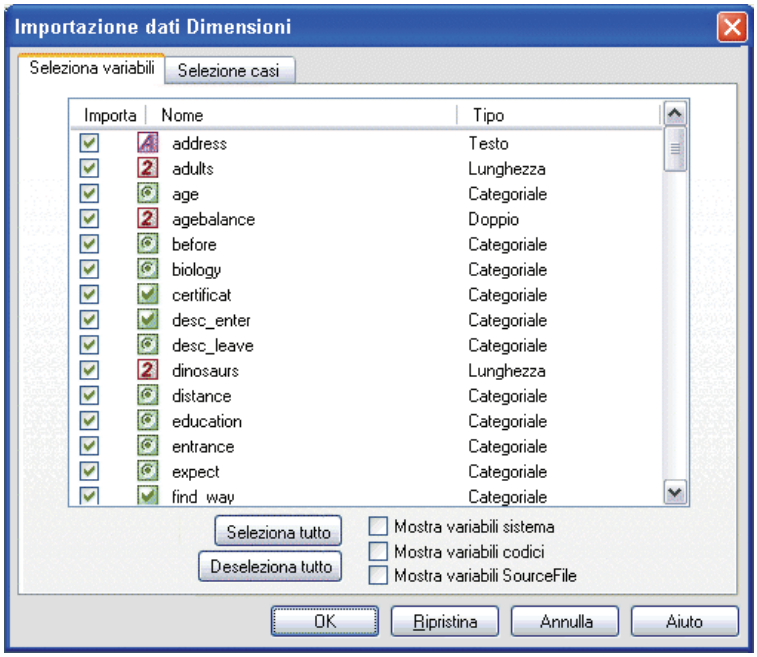
Nota: la misura in cui altre impostazioni presenti nella scheda Connessione, o qualsiasi ulteriore impostazione disponibile in altre schede della finestra di dialogo Proprietà di Data Link, possono, o meno, influire sulla lettura dei dati Dimensions in SPSS Statistics non è nota. Di conseguenza è consigliabile evitare di apportare modifiche.

Scheda Seleziona variabili

È possibile selezionare un sottoinsieme di variabili da leggere. Per impostazione predefinita, vengono visualizzate e selezionate tutte le variabili standard presenti nella sorgente di dati.

- **Mostra variabili di sistema.** Consente di visualizzare tutte le variabili di "sistema", incluse le variabili che indicano lo stato del colloquio (*in corso*, *completato*, *data di completamento* e così via). A questo punto, è possibile selezionare qualsiasi variabile di sistema che si desidera includere. Per impostazione predefinita, vengono escluse tutte le variabili di sistema.
- **Mostra variabili dei codici.** Consente di visualizzare tutte le variabili che rappresentano i codici utilizzati per le risposte aperte "Altro" per le variabili categoriali. A questo punto, è possibile selezionare qualsiasi variabile dei codici che si desidera includere. Per impostazione predefinita, vengono escluse tutte le variabili dei codici.
- **Mostra variabili SourceFile.** Consente di visualizzare tutte le variabili che contengono i nomi di file delle immagini delle risposte esaminate. A questo punto, è possibile selezionare qualsiasi variabile SourceFile che si desidera includere. Per impostazione predefinita, vengono escluse tutte le variabili SourceFile.

Figura 3-22
Importazione dati Dimensions: scheda Seleziona variabili



Scheda Selezione casi

Per le sorgenti di dati Dimensions che contengono variabili di sistema, è possibile selezionare i casi in base a numerosi criteri di tali variabili. Non è necessario includere le variabili di sistema corrispondenti nell'elenco delle variabili da leggere, ma le variabili di sistema richieste devono risultare presenti nei dati di sorgente per poter applicare i criteri di selezione. In caso contrario, i criteri di selezione corrispondenti verranno ignorati.

Stato della raccolta di dati. È possibile selezionare i dati del rispondente, i dati di test o entrambi. È inoltre possibile selezionare i casi in base a una qualsiasi combinazione dei seguenti parametri relativi allo stato dei colloqui:

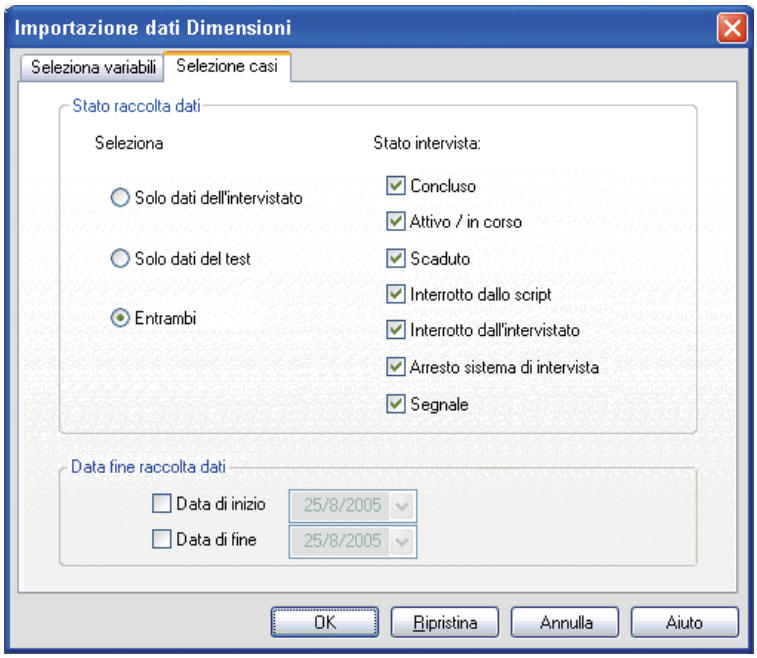
- Completato correttamente
- Attivo / in corso
- Scaduto
- Arrestato dallo script
- Arrestato dal rispondente
- Arresto del sistema di intervista
- Segnale (terminato da un'istruzione di segnale nello script)

Data di completamento della raccolta dei dati. È possibile selezionare i casi in base alla data di completamento della raccolta dei dati.

- **Data iniziale.** Vengono inclusi i casi in cui la raccolta dei dati risulta terminata alla data specificata o successivamente.

- **Data finale.** Vengono inclusi i casi in cui la raccolta dei dati risulta terminata prima della data specificata. Questo *non* comprende i casi in cui la raccolta dei dati viene completata in corrispondenza della data finale.
- Se si specifica sia una data iniziale che una finale, viene definito un intervallo di date di completamento a partire dalla data iniziale e fino a quella finale (non inclusa).

Figura 3-23
Importazione dati Dimensions: scheda Selezione casi



Informazioni sul file

Un file di dati non contiene semplicemente dati. Contiene anche informazioni sulla definizione delle variabili, ad esempio:

- Nomi delle variabili
- Formati delle variabili
- Etichette descrittive delle variabili e dei valori

Queste informazioni vengono memorizzate nella parte del file di dati riservata al dizionario. Nell'Editor dei dati è possibile visualizzare le informazioni sulla definizione delle variabili. È possibile anche visualizzare le informazioni complete sul dizionario per il file di dati attivo o qualsiasi altro file di dati.

Per visualizzare le informazioni dei file di dati

▶ Dai menu nella finestra dell'Editor dei dati, scegliere:
File
Visualizza informazioni file dati

▶ Per il file di dati attualmente aperto, scegliere File di lavoro.

▶ Per gli altri file di dati, selezionare File esterno, quindi scegliere il file di dati.

Le informazioni sul file di dati verranno visualizzate nel Viewer.

Salvataggio di file di dati

Oltre al salvataggio di file di dati in formato SPSS Statistics, è possibile salvare dati in un'ampia gamma di formati esterni, tra cui:

- Excel e altri formati di fogli elettronici
- File testo con valori delimitati da tabulazioni o virgole (CSV)
- SAS
- Stata
- Tabelle di database

Salvataggio di file di dati modificati

▶ Attivare la finestra dell'Editor dei dati facendo clic in un punto qualsiasi al suo interno.

▶ Dai menu, scegliere:

File
  Salva

Il file di dati modificato verrà salvato e la versione precedente verrà sovrascritta.

Nota: un file di dati salvato in modalità Unicode non può essere letto dalle versioni di SPSS Statistics precedenti alla 16.0. Per salvare un file di dati Unicode in un formato che può essere letto dalle versioni precedenti, aprire il file in modalità di pagina di codice e risalvarlo. Il file verrà salvato nella codifica basata sulle impostazioni internazionali correnti. Se il file contiene caratteri non riconosciuti dalle impostazioni internazionali correnti, potrebbe verificarsi una perdita di dati. Per informazioni su come passare dalla modalità Unicode alla modalità di pagina di codice, vedere Opzioni generali a pag. 522.

Salvataggio di file di dati in formati esterni

▶ Attivare la finestra dell'Editor dei dati facendo clic in un punto qualsiasi al suo interno.

▶ Dai menu, scegliere:

File
  Salva con nome...

▶ Selezionare un tipo di file dall'elenco a discesa.

▶ Specificare un nome per il nuovo file di dati.

Per scrivere i nomi delle variabili nella prima riga di file di dati in formato foglio elettronico o delimitato da tabulazioni:

▶ Fare clic su Scrivi i nomi di variabile su foglio elettronico nella finestra di dialogo Salva con nome.

Per salvare le etichette dei valori anziché i valori dei dati nei file di Excel:

- Fare clic su Salva etichette di valore se definite anziché valori di dati nella finestra di dialogo Salva con nome.

Per salvare le etichette dei valori in un file di sintassi SAS (disponibile solo se viene selezionato il tipo di file SAS).

- Fare clic su Salva etichette di valore in un file .sas nella finestra di dialogo Salva con nome.

Per informazioni sull'esportazione di dati in tabelle di database, vedere Esportazione a un database a pag. 52.

Per informazioni sull'esportazione di dati destinati ad applicazioni Dimensions, vedere Esportazione a Dimensions a pag. 64.

Salvataggio di dati: Tipi di file di dati

È possibile salvare i dati nei seguenti formati:

formato SPSS Statistics (*.sav).SPSS Statistics

- I file di dati salvati in formato SPSS Statistics non possono essere letti con versioni del prodotto precedenti alla 7.5. I file di dati salvati in modalità Unicode non possono essere letti dalle versioni di SPSS Statistics precedenti alla 16.0 Per ulteriori informazioni, vedere Opzioni generali in Capitolo 48 a pag. 522.
- Quando nella versione 10.x o 11.x si utilizzano file di dati con nomi di variabili che contengono più di otto byte, vengono utilizzate le versioni a otto byte dei nomi di variabili univoci,— mantenendo tuttavia i nomi originali delle variabili per l'utilizzo nella versione 12.0 o successiva. Nelle versioni precedenti alla 10.0, i nomi originali delle variabili più lunghi di quanto consentito andavano perduti se si salvava il file di dati.
- Se si utilizzano file di dati con variabili stringa di lunghezza superiore a 255 byte nelle versioni precedenti alla 13.0, le variabili stringa vengono divise in più variabili stringa a 255 byte.

Versione 7.0 (*.sav). Formato della versione 7.0. I file di dati salvati nel formato della versione 7.0 possono essere dalla versione 7.0 e dalle versioni precedenti, ma non includono insiemi a risposta multipla definiti né informazioni per Data Entry.

SPSS/PC+ (*.sys). Formato SPSS/PC+. Se il file di dati contiene più di 500 variabili, verranno salvate solo le prime 500. Per le variabili con più valori mancanti definiti dall'utente, i valori mancanti aggiuntivi verranno ricodificati nel primo valore mancante definito dall'utente.

SPSS Statistics Portable (*.por). Formato portatile SPSS che può essere letto da altre versioni di SPSS Statistics e dalle versioni basate su altri sistemi operativi. I nomi di variabili possono includere fino a otto byte e, se necessario, vengono convertiti automaticamente in nomi univoci a otto byte. Nella maggior parte dei casi, il salvataggio dei dati in formato portabile non è necessario, poiché i file di dati in formato SPSS Statistics dovrebbero essere indipendenti dalla piattaforma e dal sistema operativo. Non è possibile salvare i file di dati in formato portabile in modalità Unicode. Per ulteriori informazioni, vedere Opzioni generali in Capitolo 48 a pag. 522.

Delimitato da tabulazioni (*.dat). File di testo con valori delimitati da tabulazioni. (*Nota*: I caratteri di tabulazione incorporati in valori stringa sono mantenuti come tali nel file contenente valori delimitati da tabulazioni. I caratteri di tabulazione incorporati nei valori non vengono distinti da quelli che separano i valori.)

File delimitati da virgole (*.csv). File di testo con valori separati da virgola o punto e virgola. Se il separatore decimale corrente di SPSS Statistics è un punto, i valori vengono separati da virgole. Se il separatore decimale corrente di SPSS è una virgola, i valori vengono separati da punti e virgola.

ASCII fisso (*.dat). File di testo in formato fisso, in cui vengono utilizzati i formati di scrittura predefiniti per tutte le variabili. Tra i campi della variabile non sono presenti tabulazioni né spazi.

Excel 2007 (*.xlsx) Cartella di lavoro in formato XLSX di Microsoft Excel 2007. Il numero massimo di variabili è 16.000; qualsiasi variabile aggiuntiva oltre le prime 16.000 viene ignorata. Se l'insieme di dati contiene più di un milione di casi, nella cartella di lavoro vengono creati fogli multipli.

Excel da 97 a 2003 (*.xls). Cartella di lavoro di Microsoft Excel 97. Il numero massimo di variabili è 256; qualsiasi variabile aggiuntiva oltre le prime 256 viene ignorata. Se l'insieme di dati contiene più di 65.356 casi, nella cartella di lavoro vengono creati fogli multipli.

Excel 2.1 (*.xls). Foglio elettronico di Microsoft Excel 2.1. Il numero massimo di variabili è 256 e il numero massimo di righe è 16.384.

1-2-3 versione 3.0 (*.wk3). Foglio elettronico Lotus 1-2-3, versione 3,0. È possibile salvare un numero massimo di 256 variabili.

1-2-3 versione 2.0 (*.wk1). Foglio elettronico Lotus 1-2-3, versione 2,0. È possibile salvare un numero massimo di 256 variabili.

1-2-3 versione 1.0 (*.wks). Foglio elettronico Lotus 1-2-3, versione 1A. È possibile salvare un numero massimo di 256 variabili.

SYLK (*.slk). Formato Symbolic Link per fogli elettronici di Microsoft Excel e Multiplan. È possibile salvare un numero massimo di 256 variabili.

dBASE IV (*.dbf). Formato dBASE IV.

dBASE IV (*.dbf). Formato dBASE IV.

dBASE IV (*.dbf). Formato dBASE IV.

SAS v7+ Windows estensione corta (*.sd7). Versioni SAS 7–8 per Windows, formato con nome di file corto.

SAS v7+ Windows estensione lunga (*.sas7bdat). Versioni SAS 7–8 per Windows, formato con nome di file lungo.

SAS v7 per UNIX (*.ssd01). SAS v8 per UNIX.

SAS v6 per Windows (*.sd2). Formato di file SAS v6 per Windows/OS2.

SAS v6 per UNIX (*.ssd01). Formato di file SAS v6 per UNIX (Sun, HP, IBM).

SAS v6 per Alpha/OSF (*.ssd04). Formato di file SAS v6 per Alpha/OSF (DEC UNIX).

Trasporto SAS (*.xpt). File di trasporto SAS.

Stata Versione 8 Intercooled (*.dta).

Stata Versione 8 SE (*.dta).

Stata Versione 7 Intercooled (*.dta).

Stata Versione 7 SE (*.dta).

Stata Versione 6 (*.dta).

Stata Versioni 4–5 (*.dta).

Opzioni per il salvataggio di file

Nei fogli elettronici, file delimitati da tabulazioni e file delimitati da virgole, è possibile scrivere il nome delle variabili nella prima riga.

Salvataggio di file di dati in formato Excel

È possibile salvare i dati in uno dei tre formati di Microsoft Excel. Excel 2.1, Excel 97 ed Excel 2007.

- Excel 2.1 ed Excel 97 sono limitati a 256 colonne; pertanto sono incluse solo le prime 256 variabili.
- Excel 2007 è limitato a 16.000 colonne; pertanto sono incluse solo le prime 16.000 variabili.
- Excel 2.1 è limitato a 16.384 righe; pertanto sono inclusi solo i primi 16.384 casi.
- Excel 97 e Excel 2007 presentano anche limitazioni sul numero di righe per foglio, ma le cartelle di lavoro possono avere fogli multipli che vengono creati qualora venga superato il limite massimo per singolo foglio.

Tipi di variabili

Nella tabella che segue viene indicato il tipo di variabile per cui viene mantenuta la corrispondenza tra i dati originali in file SPSS Statistics e i dati esportati in file di Excel.

Tipo di variabile SPSS Statistics	Formato di dati Excel
Numerico	0.00; #,##0.00; ...
Virgola	0.00; #,##0.00; ...
Dollaro	$#,##0_); ...
Data	gg-mm-aaaa
Ora	hh:mm:ss
Stringa	Generale

Salvataggio di file di dati in formato SAS

Quando i dati vengono salvati come file in formato SAS, viene offerta una speciale gestione per vari aspetti di questi dati. Questi aspetti includono:

- Alcuni caratteri, come @, # e $, consentiti nei nomi di variabili SPSS Statistics non sono validi nei file in formato SAS. Questi caratteri vengono sostituiti con un carattere di sottolineatura durante l'esportazione dei dati.
- I nomi delle variabili SPSS Statistics che contengono caratteri con più byte (come quelli dell'alfabeto giapponese o cinese) vengono convertiti in nomi di variabili nel formato generico *Vnnn*, dove *nnn* rappresenta un numero intero.
- Le etichette delle variabili SPSS Statistics contenenti più di 40 caratteri vengono troncate se esportate in un file SAS versione 6.
- Le eventuali etichette delle variabili SPSS Statistics vengono associate alle etichette delle variabili SAS. Se i dati SPSS Statistics non includono alcuna etichetta di variabile, il nome della variabile viene associato all'etichetta della variabile SAS.
- In un file SAS è possibile includere un solo valore mancante di sistema mentre in un file SPSS Statistics possono essere inclusi più valori mancanti di sistema, di conseguenza tutti i valori mancanti di sistema in SPSS Statistics verranno associati a un singolo valore mancante di sistema nel file SAS.

Salvataggio delle etichette dei valori

Le etichette dei valori e i valori associati al file di dati possono essere salvati in un file di sintassi SAS. Ad esempio, quando vengono esportate le etichette dei valori associate al file di dati *cars.sav*, il file di sintassi generato contiene:

```
libname library '\name\' ;

proc format library = library ;
   value ORIGIN /* Paese di origine */
      1 = 'Italia'
      2 = 'Europa'
      3 = 'Giappone' ;
   value CYLINDER /* Numero di cilindri */
      3 = '3 cilindri'
      4 = '4 cilindri'
      5 = '5 cilindri'
      6 = '6 cilindri'
      8 = '8 cilindri'
   value FILTER__ /* cylrec = 1 | cylrec = 2 (FILTER) */
      0 = 'Not Selected'
      1 = 'Selected' ;
```

```
proc datasets library = library ;
modify cars;
   format    ORIGIN ORIGIN.;
   format  CYLINDER CYLINDER.;
   format  FILTER__ FILTER__.;
quit;
```

Questa caratteristica non è disponibile per il file di trasporto SAS.

Tipi di variabili

Nella tabella che segue viene indicato il tipo di variabile per cui viene mantenuta la corrispondenza tra i dati originali in file SPSS Statistics e i dati esportati in file SAS.

Tipo di variabile SPSS Statistics	Tipo di variabile SAS	Formato dati SAS.
Numerico	Numerico	12
Virgola	Numerico	12
Punto	Numerico	12
Notazione scientifica	Numerico	12
Data	Numerico	(Data) ad esempio, MMGGAA10,...
Data (ora)	Numerico	Orario 18
Dollaro	Numerico	12
Valuta personalizzata	Numerico	12
Stringa	Carattere	$8

Salvataggio di file di dati in formato Stata

- I dati possono essere scritti in formato Stata 5–8 e nei formati Intercooled e SE (solo versioni 7 e 8).
- I file di dati salvati nel formato Stata 5 possono essere letti con Stata 4.
- I primi 80 byte delle etichette delle variabili vengono salvati come etichette di variabili Stata.
- Per le variabili numeriche, i primi 80 byte delle etichette dei valori vengono salvate come etichette dei valori Stata. Per le variabili stringa, le etichette dei valori vengono eliminate.
- Per le versioni 7 e 8, i primi 32 byte dei nomi delle variabili in maiuscolo/minuscolo vengono salvate come nomi di variabili Strata. Per le versioni precedenti, i primi 8 byte dei nomi delle variabili vengono salvati come nomi di variabili Strata. Tutti i caratteri diversi da lettere, numeri e trattini di sottolineatura vengono convertiti in trattini di sottolineatura.
- I nomi delle variabili SPSS Statistics che contengono caratteri con più byte (come quelli dell'alfabeto giapponese o cinese) vengono convertiti in nomi di variabili nel formato generico *Vnnn*, dove *nnn* rappresenta un numero intero.

- Per le versioni 5–6 e le versioni Intercooled 7–8, vengono salvati i primi 80 byte dei valori stringa. Per le versioni Stata SE 7–8, vengono salvati i primi 244 byte dei valori stringa.
- Per le versioni 5–6 e le versioni Intercooled 7–8, vengono salvate solo le prime 2.047 variabili. Per le versioni Stata SE 7–8, vengono salvate solo le prime 32.767 variabili.

Tipo di variabile SPSS Statistics	Tipo di variabile Stata	Formato dati Stata.
Numerica	Numerica	g
Virgola	Numerica	g
Punto	Numerica	g
Notazione scientifica	Numerica	g
Data*, Data/ora	Numerica	D_m_Y
Ora, DOra	Numerica	g (numero di secondi)
Giorno della settimana	Numerica	g (1–7)
Mese	Numerica	g (1–12)
Dollaro	Numerica	g
Valuta personalizzata	Numerica	g
Stringa	Stringa	s

*Date, Adate, Edate, SDate, Jdate, Qyr, Moyr, Wkyr

Salvataggio di sottoinsiemi di variabili

Figura 3-24
Finestra di dialogo Salva dati come Variabili

La finestra di dialogo Salva dati come: Variabili consente di selezionare le variabili da salvare nel nuovo file di dati. Per impostazione predefinita, tutte le variabili verranno salvate. Deselezionare le variabili che non si desidera salvare oppure fare clic su Rilascia tutto e selezionare le variabili da salvare.

Solo visibili. Seleziona solo le variabili presenti negli insiemi di variabili attualmente in uso. Per ulteriori informazioni, vedere Utilizza insiemi di variabili in Capitolo 47 a pag. 518.

Per salvare un sottoinsieme di variabili

- Attivare la finestra dell'Editor dei dati facendo clic in un punto qualsiasi al suo interno.
- Dai menu, scegliere:
  File
  Salva con nome...
- Fare clic su Variabili.
- Selezionare le variabili che si desidera salvare.

Esportazione a un database

È possibile utilizzare la Procedura guidata di esportazione nel database per:

- Sostituire valori in campi di tabella di database esistenti (colonne) o aggiungere nuovi campi a una tabella.
- Aggiungere nuovi record (righe) a una tabella di database.
- Sostituire in toto una tabella di database o crearne una nuova.

Per esportare dati a un database:

- Dai menu della finestra Editor dei dati contenente i dati da esportare, scegliere:
  File
  Esporta al database
- Selezionare il database di origine.
- Per esportare i dati attenersi alle istruzioni della procedura di esportazione.

Creazione di campi di database dalle variabili SPSS Statistics

Durante la creazione di nuovi campi (aggiunta di campi a una tabella di database esistente, creazione di una nuova tabella, sostituzione di una tabella), è possibile specificare i nomi dei campi, i tipi di dati e se consentito anche la larghezza.

Nome di campo. I nomi di campo predefiniti sono uguali ai nomi di variabile SPSS Statistics. I nomi di campo possono essere sostituiti da qualsiasi nome consentito dal formato del database. Ad esempio, numerosi database consentono caratteri nei nomi di campo che non sono consentiti nei nomi di variabile, compresi gli spazi. Pertanto, un nome di variabile come *ChiamataInAttesa* può essere modificato nel nome di campo *Chiamata in attesa*.

Tipo. L'esportazione guidata attribuisce le assegnazioni iniziali del tipo di dati in base ai tipi ODBC standard o ai tipi consentiti dal formato di database selezionato più consono al formato dati SPSS Statistics definito, tenendo presente che i database possono applicare distinzioni che non hanno alcun equivalente diretto in SPSS Statistics e viceversa. Ad esempio, la maggior parte dei valori numerici in SPSS Statistics sono memorizzati sotto forma di valori con due decimali,

mentre i tipi di dati numerici del database includono numeri con decimali, interi, reali e così via. Inoltre, numerosi database non offrono equivalenti ai formati di ora di SPSS Statistics. Il tipo di dati può essere modificato in qualsiasi tipo disponibile nell'elenco a discesa.

Di norma, il tipo di dati di base (stringa o numerico) per la variabile deve corrispondere al tipo di dati di base del campo del database. La presenza di una discordanza del tipo di dati non risolta dal database genera un errore e impedisce l'esportazione dei dati nel database. Ad esempio, l'esportazione di una variabile stringa in un campo di database di tipo numerico genera un errore in presenza di valori della variabile stringa contenenti caratteri non numerici.

Larghezza. È possibile modificare la larghezza definita del tipo di campo (char, varchar) stringa. La larghezza dei campi numerici è definita dal tipo di dati.

Per impostazione predefinita, i formati di variabile SPSS Statistics vengono associati ai tipi di campo del database sulla base dello schema generale seguente. I tipi di campo del database variano a seconda del database.

Formato delle variabili SPSS Statistics	Tipo di campo del database
Numerica	Decimale o doppio
Virgola	Decimale o doppio
Punto	Decimale o doppio
Notazione scientifica	Decimale o doppio
Data	Data o data/ora o timestamp
Data e ora	Data/ora o timestamp
Ora, DOra	Decimale o doppio (numero di secondi)
Giorno della settimana	Intero (1–7)
Mese	Intero (1–12)
Dollaro	Decimale o doppio
Valuta personalizzata	Decimale o doppio
Stringa	Char o Varchar

Valori mancanti definiti dall'utente

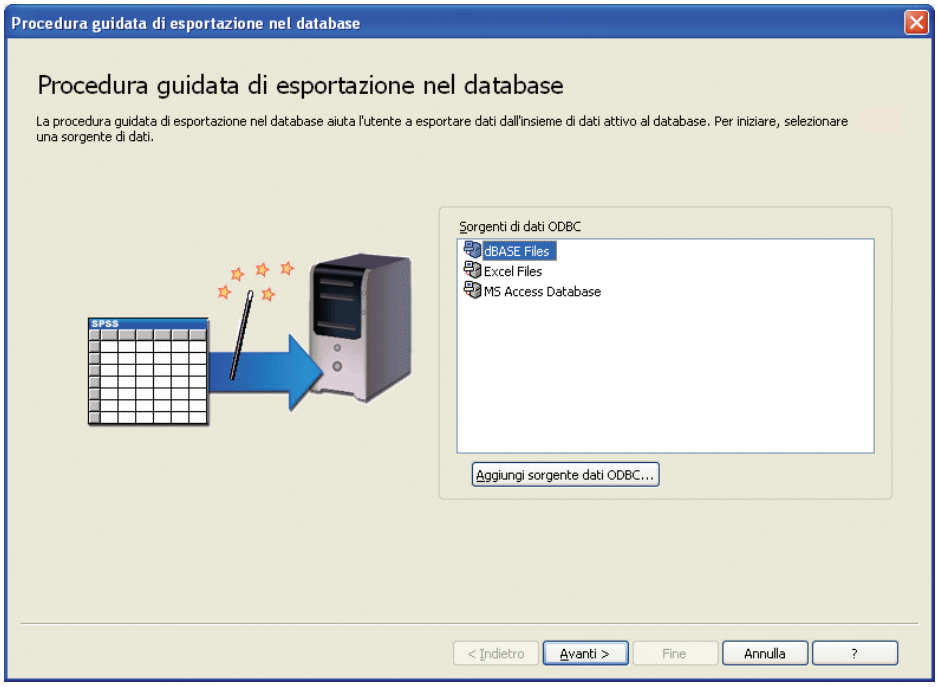
Durante l'esportazione di dati da variabili a campi di database, sono disponibili due opzioni per l'elaborazione dei valori definiti dall'utente:

- **Esporta come valori validi.** I valori mancanti definiti dall'utente vengono considerati come valori standard, validi non mancanti.
- **Esporta valori numerici mancanti definiti dall'utente come nulli ed esporta valori stringa mancanti definiti dall'utente come spazi vuoti.** I valori numerici mancanti definiti dall'utente sono considerati come valori mancanti di sistema. I valori stringa mancanti definiti dall'utente vengono convertiti in spazi vuoti; le stringhe non possono essere valori mancanti di sistema.

Selezione di una sorgente dati

Nel primo riquadro della Procedura guidata di esportazione nel database, selezionare la sorgente dati a cui esportare i dati.

Figura 3-25
Procedura guidata di esportazione nel database; selezione di una sorgente dati


Procedura guidata di esportazione nel database
Procedura guidata di esportazione nel database
La procedura guidata di esportazione nel database aiuta l'utente a esportare dati dall'insieme di dati attivo al database. Per iniziare, selezionare una sorgente di dati.
Sorgenti di dati ODBC
dBASE Files
Excel Files
MS Access Database
Aggiungi sorgente dati ODBC...
< Indietro
Avanti >
Fine
Annulla
?

È possibile esportare dati verso qualsiasi sorgente database per la quale si dispone del driver ODBC appropriato. (*Nota*: l'esportazione di dati verso sorgenti dati OLE DB non è supportata.)

Se non è stata configurata alcuna sorgente dati ODBC o se si desidera aggiungerne una nuova, fare clic su Aggiungi sorgente dati ODBC.

- Nei sistemi operativi Linux, questo pulsante non è disponibile. Le sorgenti dati ODBC vengono specificate nel file *odbc.ini* e le variabili di ambiente *ODBCINI* devono essere impostate sulla posizione di tale file. Per ulteriori informazioni, vedere la documentazione relativa ai driver del database.
- Nella modalità di analisi distribuita (disponibile con SPSS Statistics Server) questo pulsante non è disponibile. Per aggiungere sorgenti dati nella modalità di analisi distribuita, rivolgersi all'amministratore di sistema.

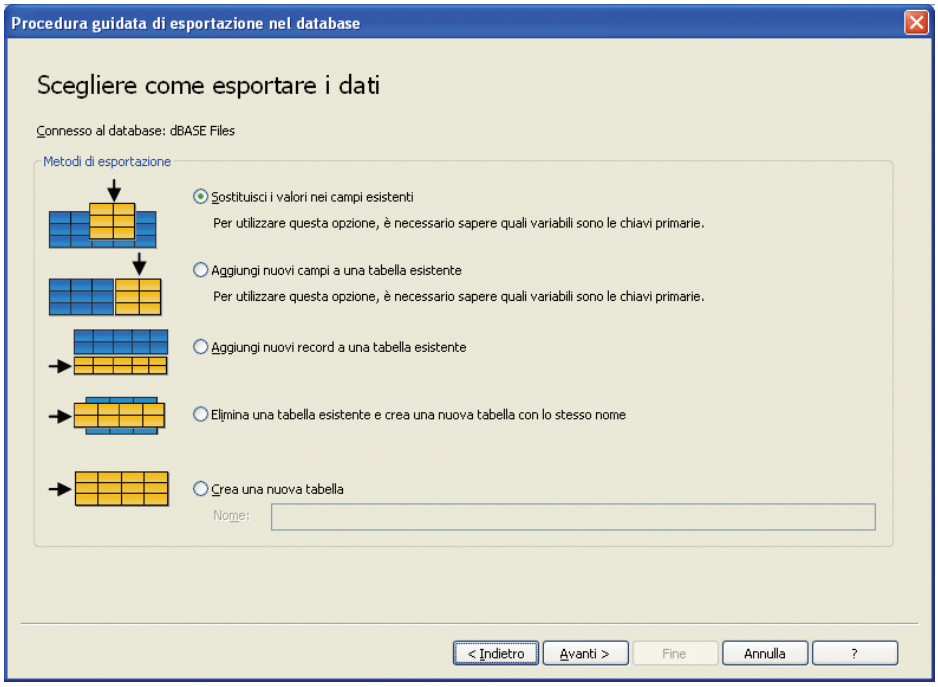
Una sorgente dati ODBC include due informazioni fondamentali: il driver che verrà utilizzato per accedere ai dati e la posizione del database a cui si desidera accedere. Per specificare le sorgenti dati, è necessario che siano installati i driver appropriati Per la modalità di analisi locale, è possibile installare i driver per vari formati di database dal CD di installazione di SPSS Statistics.

Alcune sorgenti dati richiedono un ID e password di accesso per il passaggio alla fase successiva.

Scelta della modalità di esportazione dei dati

Dopo la selezione della sorgente dati, indicare il modo in cui esportare i dati.

Figura 3-26
Procedura guidata di esportazione nel database; selezione della modalità di esportazione


Procedura guidata di esportazione nel database
Scegliere come esportare i dati
Connesso al database: dBASE Files
Metodi di esportazione
Sostituisci i valori nei campi esistenti
Per utilizzare questa opzione, è necessario sapere quali variabili sono le chiavi primarie.
Aggiungi nuovi campi a una tabella esistente
Per utilizzare questa opzione, è necessario sapere quali variabili sono le chiavi primarie.
Aggiungi nuovi record a una tabella esistente
Elimina una tabella esistente e crea una nuova tabella con lo stesso nome
Crea una nuova tabella
Nome:
< Indietro
Avanti >
Fine
Annulla
?

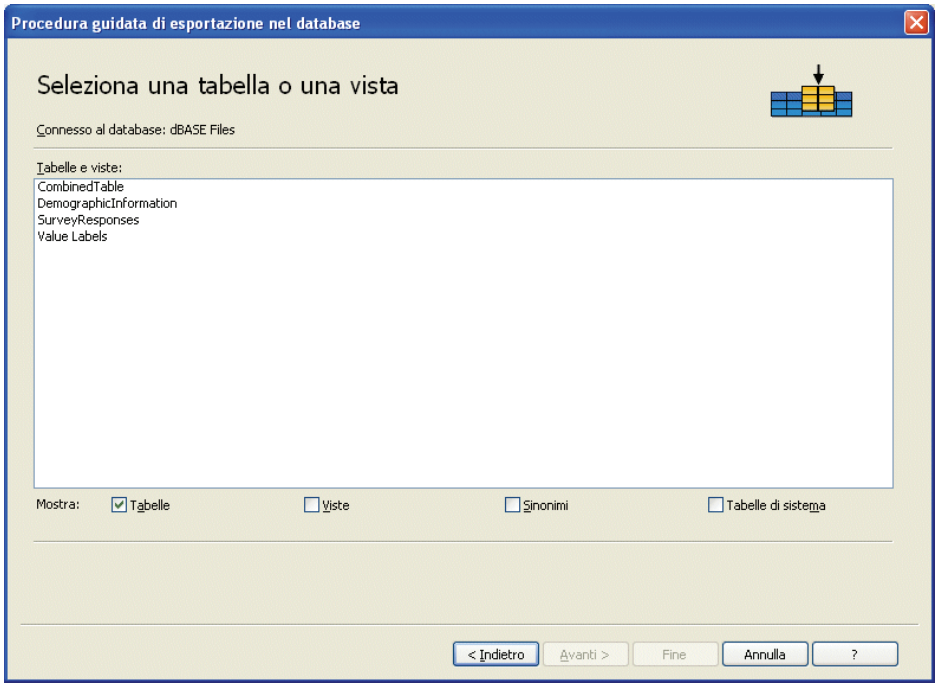
Per l'esportazione di dati verso un database, sono disponibili le seguenti opzioni:

- **Sostituisci i valori nei campi esistenti.** Sostituisce in una tabella esistente i valori dei campi selezionati con valori di variabili selezionate nell'insieme di dati attivo. Per ulteriori informazioni, vedere Sostituzione dei valori nei campi esistenti a pag. 58.
- **Aggiungi nuovi campi a una tabella esistente.** Crea nuovi campi in una tabella esistente contenente valori di variabili selezionate nell'insieme di dati attivo. Per ulteriori informazioni, vedere Aggiunta di nuovi campi a pag. 59. Questa opzione non è disponibile per i file Excel.
- **Aggiungi nuovi record a una tabella esistente.** Aggiunge a una tabella esistente nuovi record (righe) contenenti valori da casi nell'insieme di dati attivo. Per ulteriori informazioni, vedere Aggiunta di nuovi record (casi) a pag. 60.
- **Elimina una tabella esistente e crea una nuova tabella con lo stesso nome.** Elimina la tabella specificata e ne crea una nuova con lo stesso nome contenente variabili selezionate nell'insieme di dati attivo. Tutte le informazioni della tabella di origine, incluse le definizioni delle proprietà dei campi (ad esempio, chiavi primarie, tipi di dati) vengono perse. Per ulteriori informazioni, vedere Creazione di una nuova tabella o sostituzione di una tabella a pag. 62.
- **Crea una nuova tabella.** Crea nel database una nuova tabella contenente dati dalle variabili selezionate nell'insieme di dati attivo. Sono validi tutti i nomi consentiti per la tabella dalla sorgente dati. Non è consentito duplicare il nome di una tabella o vista esistente nel database. Per ulteriori informazioni, vedere Creazione di una nuova tabella o sostituzione di una tabella a pag. 62.

Selezione della tabella

L'operazione di modifica o sostituzione di una tabella nel database comporta la preliminare selezione della tabella. Questo riquadro della Procedura guidata di esportazione nel database visualizza un elenco di tabella e viste nel database selezionato.

Figura 3-27
Procedura guidata di esportazione nel database; selezione della tabella o della vista



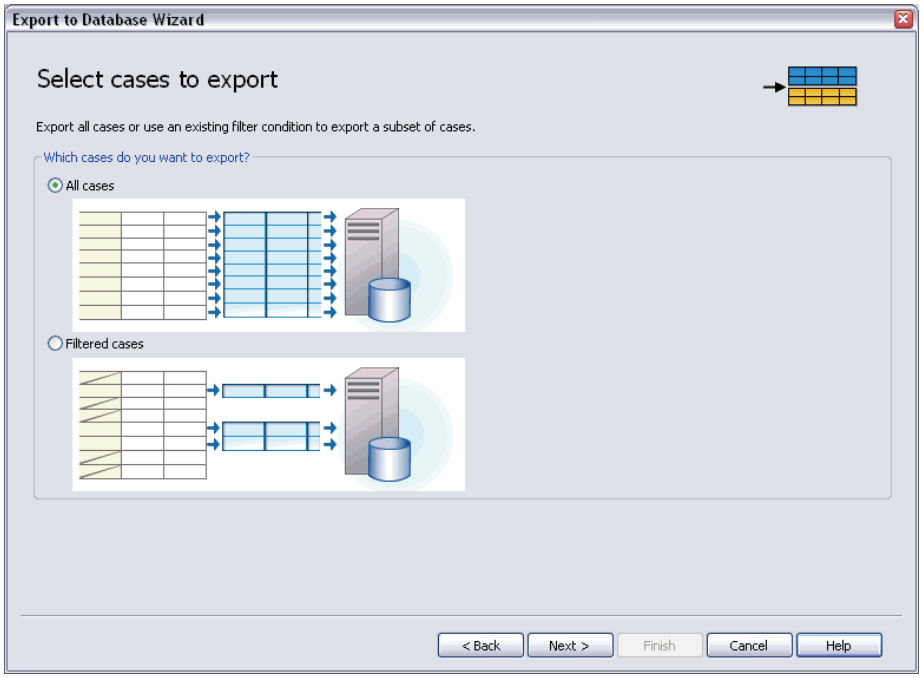
Per impostazione predefinita, l'elenco visualizza solo le tabelle di database standard. È possibile impostare il tipo di oggetti da visualizzare nell'elenco:

- **Tabelle.** Tabelle di database standard.
- **Viste.** Le viste sono "tabelle" virtuali o dinamiche definite da query. Possono includere di più tabelle e/o campi frutto di calcoli basati sui valori di altri campi. Le viste consentono, non senza vincoli a seconda della relativa struttura, l'aggiunta di record o la sostituzione di valori di campi esistenti. Ad esempio, non è consentito modificare un campo derivato, aggiungere campi a una vista o sostituire una vista.
- **Sinonimi.** Un sinonimo è l'alias di una tabella o di una vista, di norma definito in una query.
- **Tabelle di sistema.** Le tabelle di sistema definiscono le proprietà del database. In alcuni casi, le tabelle di database standard sono classificate come tabelle di sistema e possono essere visualizzate solo selezionando questa opzione. Spesso solo agli amministratori possono accedere alle vere tabelle di sistema.

Selezione dei casi da esportare

La selezione dei casi nella Procedura guidata di esportazione nel database offre due opzioni: selezione di tutti i casi o selezione dei casi filtrati in base a criteri definiti in precedenza. In assenza di filtro dei casi, questo riquadro non viene visualizzato e tutti i casi dell'insieme di dati attivo vengono esportati.

Figura 3-28
Procedura guidata di esportazione nel database; selezione dei casi da esportare


Export to Database Wizard
Select cases to export
Export all cases or use an existing filter condition to export a subset of cases.
Which cases do you want to export?
All cases
Filtered cases
< Back
Next >
Finish
Cancel
Help

Per informazioni sulla definizione di una condizione di filtro per la selezione dei casi, vedere Seleziona casi a pag. 193.

Abbinamento di casi e record

Per l'aggiunta di campi (colonne) a una tabella esistente o per la sostituzione di valori di campi esistenti, è necessario accertarsi che ciascun caso (riga) nell'insieme di dati attivo sia correttamente abbinato al record corrispondente nel database.

- Nel database, il campo o insieme di campi che identifica in modo univoco ciascun record è noto anche con il nome di **chiave primaria**.
- È necessario identificare la o le variabili che corrispondono al campo o ai campi di chiave primaria o gli altri campi che identificano in modo univoco ciascun record.
- I campi non devono per forza corrispondere alla chiave primaria nel database, ma il valore del campo o la combinazione dei valori del campo deve essere univoca per ciascun caso.

Per far corrispondere le variabili ai campi nel database che identificano in modo univoco ciascun record:

- Trascinare le variabili sui campi corrispondenti del database.

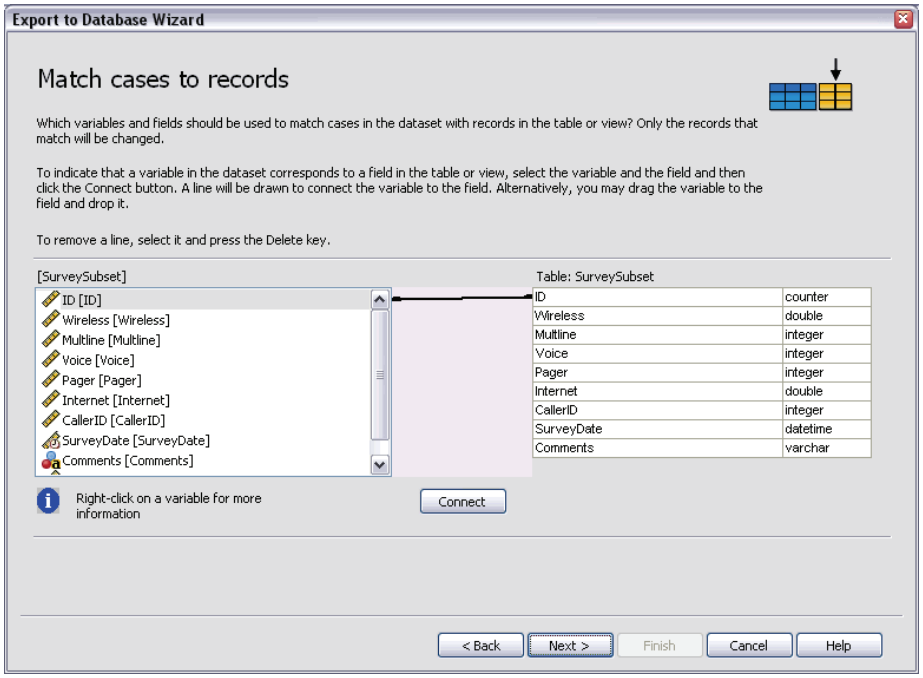
o un

- Selezionare una variabile dall'elenco di variabili, selezionare il campo corrispondente nella tabella di database, quindi fare clic su Connetti.

Per eliminare una connessione:

- Selezionare la connessione e premere il tasto Canc.

Figura 3-29
Procedura guidata di esportazione nel database; abbinamento di casi e record


Export to Database Wizard
Match cases to records
Which variables and fields should be used to match cases in the dataset with records in the table or view? Only the records that match will be changed.
To indicate that a variable in the dataset corresponds to a field in the table or view, select the variable and the field and then click the Connect button. A line will be drawn to connect the variable to the field. Alternatively, you may drag the variable to the field and drop it.
To remove a line, select it and press the Delete key.
[SurveySubset]
ID [ID]
Wireless [Wireless]
Multline [Multline]
Voice [Voice]
Pager [Pager]
Internet [Internet]
CallerID [CallerID]
SurveyDate [SurveyDate]
Comments [Comments]
Table: SurveySubset
ID counter
Wireless double
Multline integer
Voice integer
Pager integer
Internet double
CallerID integer
SurveyDate datetime
Comments varchar
Right-click on a variable for more information
Connect
< Back
Next >
Finish
Cancel
Help

Nota: I nomi delle variabili e i nomi dei campi del database non sono sempre identici (dal momento che i nomi dei campi del database possono contenere caratteri non consentiti nei nomi delle variabili SPSS Statistics), ma se l'insieme di dati attivo è stato creato dalla tabella di database in corso di modifica, è probabile che i nomi o le etichette delle variabili siano almeno simili ai nomi dei campi del database.

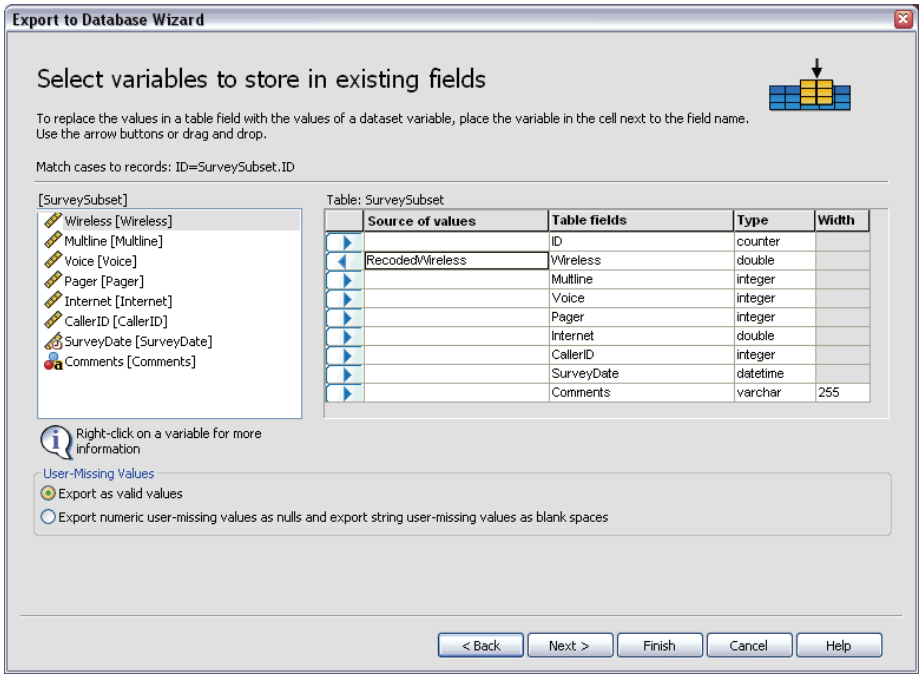
Sostituzione dei valori nei campi esistenti

Per sostituire i valori dei campi esistenti in un database:

- Nel riquadro Scegliere come esportare i dati della Procedura guidata di esportazione nel database, selezionare Sostituisci i valori nei campi esistenti.
- Nel riquadro Seleziona una tabella o una vista, selezionare la tabella di database.

- Nel riquadro Confrontare i casi con i record, abbinare le variabili univoche per ciascun caso ai corrispondenti nomi di campo nel database.
- Per ciascun campo i cui valori sono da sostituire, trascinare la variabile contenente i nuovi valori nella colonna Sorgente dei valori, a fianco del nome di campo corrispondente nel database.

Figura 3-30
Procedura guidata di esportazione nel database; sostituzione di valori in campi esistenti



  - Di norma, il tipo di dati di base (stringa o numerico) per la variabile deve corrispondere al tipo di dati di base del campo del database. La presenza di una discordanza del tipo di dati non risolta dal database genera un errore e impedisce l'esportazione dei dati nel database. Ad esempio, l'esportazione di una variabile stringa in un campo di database di tipo numerico (ad esempio doppio, reale, intero) genera un errore in presenza di valori della variabile stringa contenenti caratteri non numerici. La lettera *a* nell'icona a fianco di una variabile indica una variabile stringa.
  - Non è consentito modificare il nome, tipo o larghezza del campo. Gli attributi originali del campo del database vengono conservati; solo i valori vengono sostituiti.

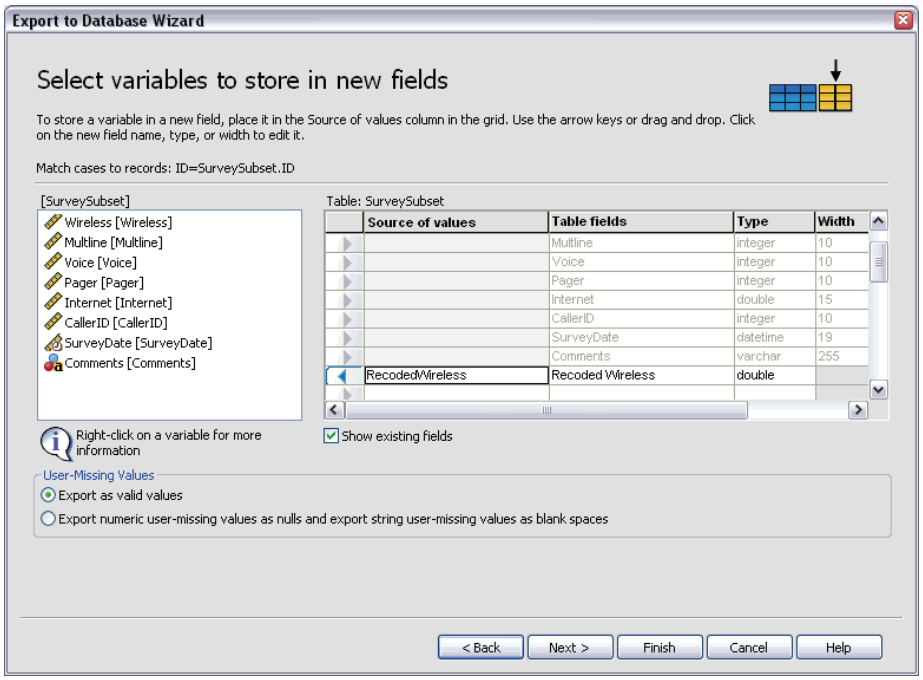
Aggiunta di nuovi campi

Per aggiungere nuovi campi a una tabella di database esistente:

- Nel riquadro Scegliere come esportare i dati della Procedura guidata di esportazione nel database, selezionare Aggiungi nuovi campi a una tabella esistente.
- Nel riquadro Seleziona una tabella o una vista, selezionare la tabella di database.
- Nel riquadro Confrontare i casi con i record, abbinare le variabili univoche per ciascun caso ai corrispondenti nomi di campo nel database.

- Trascinare le variabili da aggiungere come nuovi campi alla colonna Sorgente dei valori.

Figura 3-31
Procedura guidata di esportazione nel database; aggiunta di nuovi campi a una tabella esistente



Per informazioni sui nomi di campo e i tipi di dati, vedere la sezione sulla creazione di campi di database da SPSS Statistics variabili in Esportazione a un database a pag. 52.

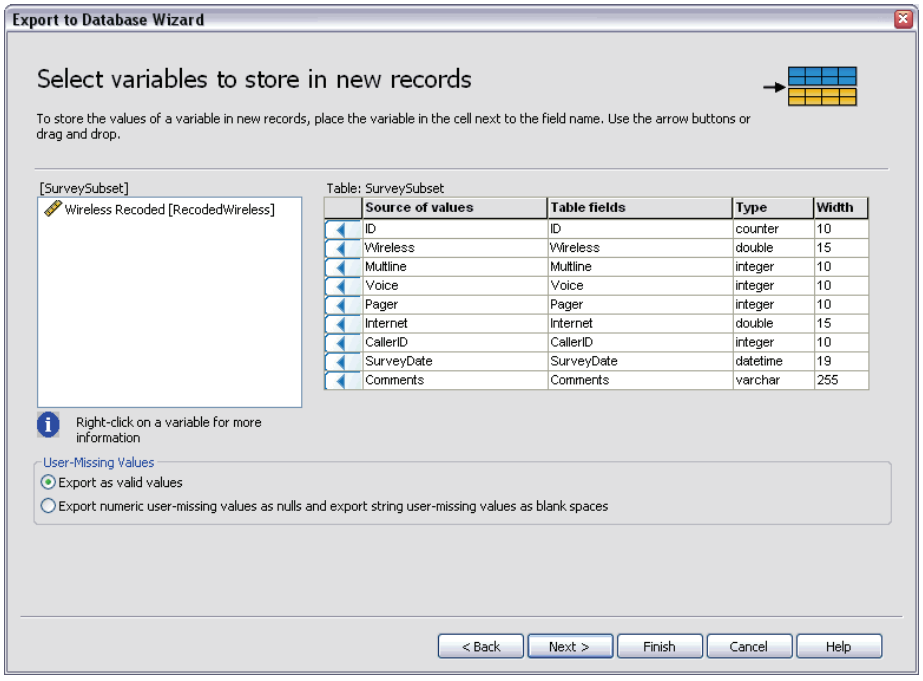
Mostra campi esistenti. Selezionare questa opzione per visualizzare un elenco di campi esistenti. Questo riquadro non può essere usato per sostituire i campi esistenti ma è utile per conoscere i campi già presenti in tabella. Per sostituire i valori di campi esistenti, vedere Sostituzione dei valori nei campi esistenti a pag. 58.

Aggiunta di nuovi record (casi)

Per aggiungere nuovi record (righe) a una tabella di database:

- Nel riquadro Scegliere come esportare i dati della Procedura guidata di esportazione nel database, selezionare Aggiungi nuovi record a una tabella esistente.
- Nel riquadro Seleziona una tabella o una vista, selezionare la tabella di database.
- Per abbinare le variabili dell'insieme di dati attivo ai campi della tabella, trascinare le variabili nella colonna Sorgente dei valori.

Figura 3-32
Procedura guidata di esportazione nel database; aggiunta di record (casi) a una tabella


Export to Database Wizard
Select variables to store in new records
To store the values of a variable in new records, place the variable in the cell next to the field name. Use the arrow buttons or drag and drop.
[SurveySubset]
Wireless Recoded [RecodedWireless]
Table: SurveySubset
Source of values
Table fields
Type
Width
ID ID counter 10
Wireless Wireless double 15
Multline Multline integer 10
Voice Voice integer 10
Pager Pager integer 10
Internet Internet double 15
CallerID CallerID integer 10
SurveyDate SurveyDate datetime 19
Comments Comments varchar 255
Right-click on a variable for more information
User-Missing Values
Export as valid values
Export numeric user-missing values as nulls and export string user-missing values as blank spaces
< Back
Next >
Finish
Cancel
Help

La procedura selezionerà automaticamente tutte le variabili corrispondenti ai campi esistenti in base alle informazioni sulla tabella di database originale memorizzate nell'insieme di dati attivo (se disponibile) e/o sui nomi di variabile identici ai nomi di campo. Questo abbinamento iniziale automatico rappresenta solo un'indicazione e non impedisce la modifica del modo in cui le variabili vengono abbinate ai campi del database.

Durante l'aggiunta di nuovi record a una tabella esistente, attenersi alle regole/vincoli di base seguenti:

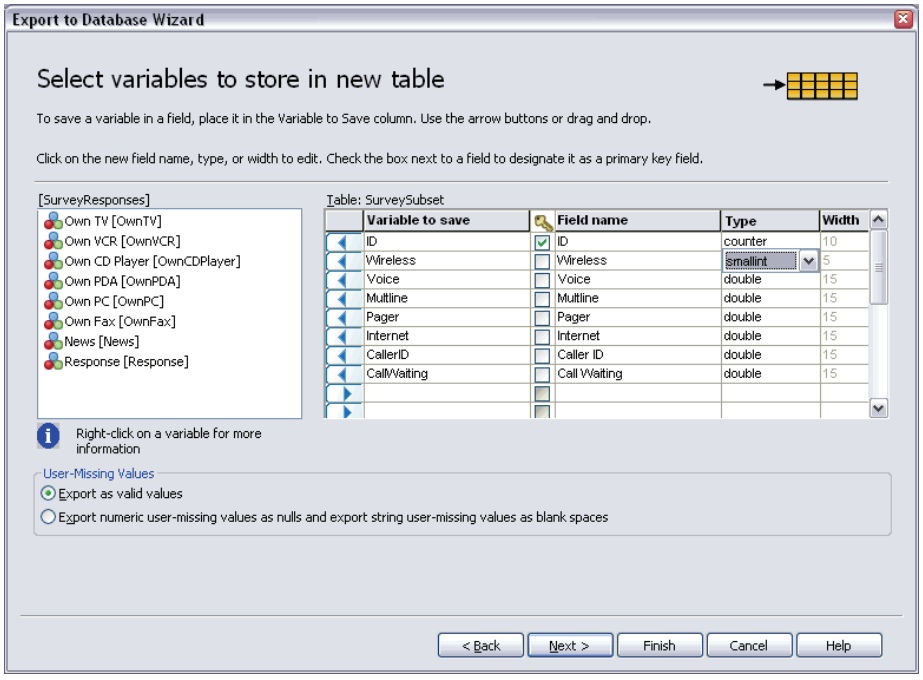
- Tutti i casi (o tutti i casi selezionati) dell'insieme di dati attivo vengono aggiunti alla tabella. Se uno dei casi è il duplicato di un record esistente nel database, è possibile che venga prodotto un errore in presenza di una chiave duplicata. Per informazioni sull'esportazione dei soli casi selezionati, vedere Selezione dei casi da esportare a pag. 57.
- È possibile usare i valori di nuove variabili create nella sessione come i valori dei campi esistenti; tuttavia, non è possibile aggiungere nuovi campi o modificare i nomi di campi esistenti. Per aggiungere nuovi campi a una tabella di database, vedere Aggiunta di nuovi campi a pag. 59.
- Tutti i campi di database esclusi o quelli non abbinati a una variabile saranno privi di valori a fronte dei record aggiunti nella tabella di database. (Se la cella Sorgente dei valori è vuota, significa che non esiste una variabile abbinata al campo.)

Creazione di una nuova tabella o sostituzione di una tabella

Per creare una nuova tabella di database o sostituire una tabella di database esistente:

- Nel riquadro Scegliere come esportare i dati della procedura di esportazione, selezionare Elimina una tabella esistente e crea una nuova tabella con lo stesso nome o Crea una nuova tabella e digitare un nome per la nuova variabile.
- Se si tratta di sostituzione di una tabelle esistente, nel riquadro Seleziona una tabella o una vista, selezionare la tabella di database.
- Trascinare le variabili nella colonna Variabile da salvare.
- Se necessario, è possibile designare variabili/campi che definiscono la chiave primaria nonché modificare i nomi dei campi e il tipo di dati.

Figura 3-33
Procedura guidata di esportazione nel database; selezione delle variabili per una nuova tabella



Chiave primaria. Per designare variabili come chiave primaria nella tabella del database, selezionare nella colonna la casella associata all'icona chiave.

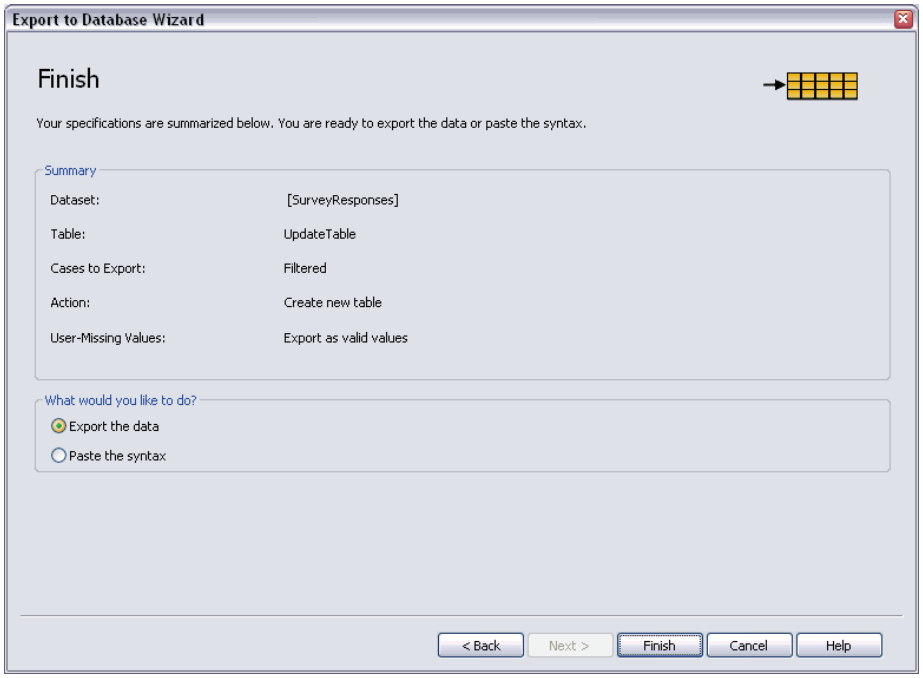
- Tutti i valori della chiave primaria devono essere univoci, altrimenti viene generato un errore.
- Se una sola variabile è chiave primaria, tutti i record (casi) devono presentare un valore univoco per quella variabile.
- Se più variabili sono chiave primaria, si tratterà di chiave primaria composita e la *combinazione* di valori per le variabili selezionate deve essere univoca per ciascun caso.

Per informazioni sui nomi di campo e i tipi di dati, vedere la sezione sulla creazione di campi di database da SPSS Statistics variabili in Esportazione a un database a pag. 52.

Completamento della Procedura guidata di esportazione nel database

L'ultimo riquadro della Procedura guidata di esportazione nel database presenta un riepilogo che indica i dati che verranno esportati nonché la modalità di esportazione. Viene inoltre offerta la possibilità di scegliere se esportare i dati o se incollare la sintassi dei comandi pertinenti nell'apposita finestra.

Figura 3-34
Procedura guidata di esportazione nel database; riquadro conclusivo



Informazioni di riepilogo

- **Insieme di dati.** Nome della sessione SPSS Statistics per l'insieme di dati che verrà usato per l'esportazione dei dati. Tale informazione è particolarmente utile in presenza di più sorgenti dati aperte. Le sorgenti dati aperte tramite l'interfaccia utente grafica (ad esempio Creazione guidata di query), vengono automaticamente associate a nomi di tipo *InsiemeDati1*, *InsiemeDati2* e così via. Una sorgente dati aperta tramite la sintassi dei comandi risulterà associata a un nome di insieme dati solo se questo viene esplicitamente assegnato.
- **Tabella.** Nome della tabella da modificare o creare.
- **Casi da esportare.** Esportazione di tutti i casi o solo di quelli definiti da un filtro. Per ulteriori informazioni, vedere Selezione dei casi da esportare a pag. 57.
- **Azione.** Indica il tipo di modifica applicata al database (ad esempio, creazione di una nuova tabella, aggiunta di campi o record a una tabella esistente).
- **Valori mancanti definiti dall'utente.** I valori mancanti definiti dall'utente possono essere esportati come valori validi o considerati come valori di sistema mancanti per le variabili numeriche e convertiti in spazi vuoti per le variabili stringa. Questa impostazione è disponibile nel riquadro di selezione delle variabili da esportare.

Esportazione a Dimensions

La finestra di dialogo Esporta a Dimensions crea un file di dati di SPSS Statistics e un file di metadati Dimensions che può essere usato per leggere i dati nelle applicazioni Dimensions come mrInterview e mrTables. Ciò è particolarmente utile per lo scambio di dati tra le applicazioni SPSS Statistics e Dimensions. Ad esempio, è possibile leggere una sorgente dati mrInterview in SPSS Statistics, calcolare nuove variabili e salvare i dati in un modulo che può essere letto da mrTables, senza perdita degli attributi dei metadati di origine.

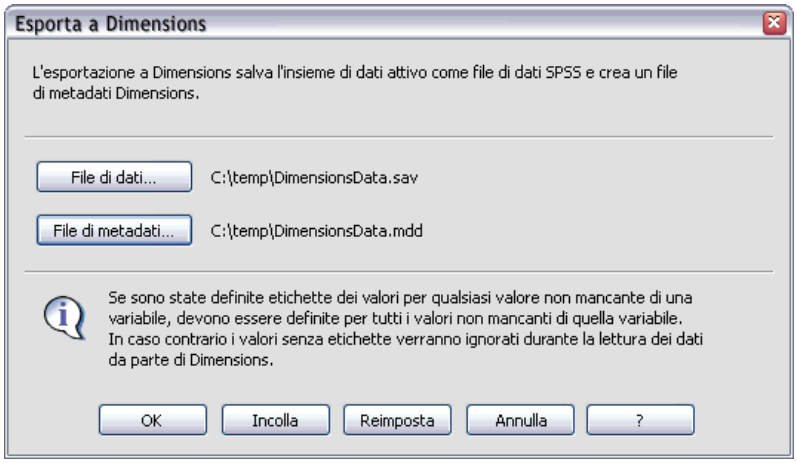
Per esportare dati leggibili nelle applicazioni Dimensions:

▶ Dai menu della finestra Editor dei dati contenente i dati da esportare, scegliere:
File
Esporta in Dimensions

▶ Fare clic su File di dati per specificare il nome e la posizione del file di dati in formato SPSS Statistics.

▶ Fare clic su File di metadati per specificare il nome e la posizione del file di metadati Dimensions.

Figura 3-35
Finestra di dialogo Esporta a Dimensions


Esporta a Dimensions
L'esportazione a Dimensions salva l'insieme di dati attivo come file di dati SPSS e crea un file di metadati Dimensions.
File di dati...
C:\temp\DimensionsData.sav
File di metadati...
C:\temp\DimensionsData.mdd
Se sono state definite etichette dei valori per qualsiasi valore non mancante di una variabile, devono essere definite per tutti i valori non mancanti di quella variabile. In caso contrario i valori senza etichette verranno ignorati durante la lettura dei dati da parte di Dimensions.
OK
Incolla
Reimposta
Annulla
?

Per le variabili e gli insiemi di dati nuovi non creati da sorgenti dati Dimensions, gli attributi di variabile SPSS Statistics sono associati agli attributi dei metadati Dimensions nel file dei metadati in base ai metodi descritti nella documentazione SAV DSC disponibile in Dimensions Development Library.

Se l'insieme di dati attivo è stato creato da una sorgente dati Dimensions:

- Il nuovo file di metadati viene creato unendo gli attributi dei metadati di origine con gli attributi dei metadati della nuova variabile, aggiungendovi tutte le modifiche apportate alle variabili di origine che possono incidere sugli attributi dei metadati (ad esempio aggiunte o modifiche alle etichette dei valori).
- Riguardo le variabili di origine lette dalla sorgente dati Dimensions, tutti gli attributi di metadati non riconosciuti da SPSS Statistics vengono conservati nello stato originale. Ad esempio, SPSS Statistics converte le variabili di griglia in variabili SPSS Statistics standard,

ma i metadati che definiscono tali variabili di griglia vengono conservati con il salvataggio del nuovo file di metadati.

- In presenza di variabili Dimensions automaticamente rinominate per conformarle alle regole SPSS Statistics di denominazione delle variabili, nel file dei metadati i nomi delle variabili vengono riportati ai valori originali di Dimensions.

La presenza o l'assenza di etichette di valori può incidere sugli attributi dei metadati delle variabili e di conseguenza sul modo in cui tali variabili vengono lette dalle applicazioni Dimensions. Se sono state definite delle etichette di valori per valori di una variabile non mancanti, queste devono essere definite per tutti i valori non mancanti di quella variabile; altrimenti i valori senza etichetta verranno eliminati alla lettura del file di dati da Dimensions.

Protezione dei dati originali

Per evitare che i dati originali vengano modificati o eliminati accidentalmente, è possibile contrassegnare il file come di sola lettura.

▶ Dai menu dell'Editor dei dati, scegliere:
File
Contrassegna file in sola lettura

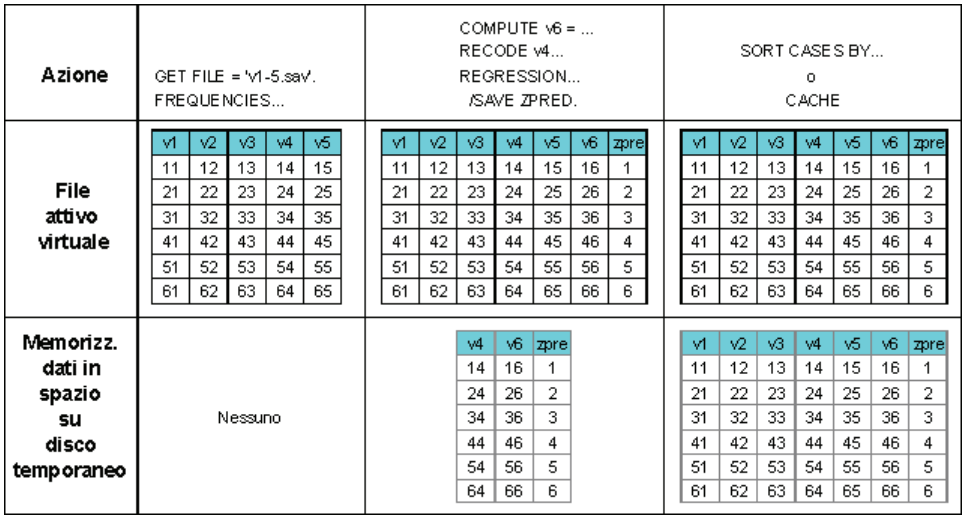
Se in seguito si apportano modifiche ai dati e quindi si prova a salvare il file, sarà possibile salvarlo solo con un nome file diverso. In questo modo, i dati originali non verranno alterati.

È possibile reimpostare i permessi di lettura/scrittura per il file scegliendo Contrassegna file come di scrittura dal menu File.

File attivo virtuale

Il file attivo virtuale consente di lavorare con file di dati di grandi dimensioni senza che sia necessario liberare una quantità uguale o maggiore di spazio su disco temporaneo. Nella maggior parte delle procedure di analisi e rappresentazione grafica, i dati di origine vengono riletti ogni volta che si esegue una procedura diversa. Le procedure che modificano i dati necessitano di una certa quantità di spazio su disco temporaneo per tenere traccia delle modifiche. Inoltre, alcune azioni richiedono sempre una quantità di spazio su disco sufficiente a contenere almeno una copia dell'intero file di dati.

Figura 3-36
Requisiti di spazio su disco temporaneo



Le azioni che non richiedono spazio su disco temporaneo sono:

- Lettura dei file di dati di SPSS Statistics
- Unione di due o più file di dati di SPSS Statistics
- Lettura di tabelle di database con la procedura Creazione guidata di query
- Unione di file di dati di SPSS Statistics con tabelle di database
- Esecuzione delle procedure di lettura dei dati (ad esempio, Frequenze, Tavole di contingenza, Esplora)

Le azioni che creano una o più colonne di dati nello spazio su disco temporaneo sono:

- Calcolo di nuove variabili
- Registrazione delle variabili esistenti
- Esecuzione di procedure che creano o modificano variabili (ad esempio, il salvataggio dei valori attesi nella regressione lineare)

Le azioni che creano una copia dell'intero file di dati nello spazio su disco temporaneo sono:

- Lettura di file di Excel
- Esecuzione di procedure che ordinano i dati (ad esempio, Ordina casi, Dividi)
- Lettura dei dati con i comandi `GET TRANSLATE` o `DATA LIST`
- Uso dell'utilità di inserimento dati nella cache o del comando `CACHE`
- Avvio di altre applicazioni da SPSS Statistics che leggono il file di dati (ad esempio, AnswerTree, DecisionTime)

Nota: il comando `GET DATA` offre funzionalità simili a quelle offerte dal comando `DATA LIST`, senza però creare una copia dell'intero file di dati nello spazio su disco temporaneo. Il comando `SPLIT FILE` nella sintassi dei comandi non esegue l'ordinamento del file di dati e quindi

non crea una copia del file di dati. Questo comando richiede tuttavia i dati ordinati per un funzionamento appropriato e l'interfaccia della finestra di dialogo relativa a questa procedura ordina automaticamente il file di dati, creando una copia dell'intero file di dati. La sintassi dei comandi non è disponibile nella versione per studenti.

Azioni che creano una copia intera del file di dati per impostazione predefinita:

- Lettura di database con Creazione guidata di query.
- Lettura di file di testo con Importazione guidata di testo

L'Importazione guidata testo fornisce un'impostazione opzionale che consente di memorizzare automaticamente i dati nella cache. Questa opzione è selezionata per impostazione predefinita. È possibile disattivarla deselezionando la casella di controllo Copia i dati in locale. Per la Creazione guidata di query, è possibile incollare la sintassi dei comandi generata ed eliminare il comando `CACHE`.

Creazione di una copia dei dati

Benché il file attivo virtuale consenta di ridurre notevolmente la quantità di spazio su disco temporaneo richiesta, l'assenza di una copia temporanea del file "attivo" comporta la rilettura dei dati di origine ogni volta che si esegue una procedura. Per i file di dati di grandi dimensioni letti da una sorgente esterna, la creazione di una copia temporanea dei dati può migliorare le prestazioni. Nel caso di tabelle lette da un database, ciò significa che la query SQL che legge i dati dal database deve essere rieseguita per ogni comando o procedura che richiede la lettura dei dati. Dato che praticamente tutte le procedure di analisi statistica e di rappresentazione grafica richiedono una lettura dei dati, la query SQL viene rieseguita a ogni procedura, provocando un notevole aumento dei tempi di elaborazione quando si esegue un numero significativo di procedure.

Se il computer su cui viene eseguita l'analisi (sia esso un computer locale o un server remoto) dispone di spazio su disco sufficiente, è possibile evitare l'esecuzione continua delle query SQL e ridurre i tempi di elaborazione creando una copia cache dei dati del file attivo. Tale copia è una copia temporanea completa dei dati.

Nota: per impostazione predefinita, Creazione guidata di query crea automaticamente una copia di dati ma se si utilizza il comando `GET DATA` nella sintassi dei comandi per leggere un database, tale copia non viene creata automaticamente. La sintassi dei comandi non è disponibile nella versione per studenti.

Per creare una copia dei dati

▶ Dai menu, scegliere:
File
Copia dati...

▶ Fare clic su OK o su In cache ora.

Se si fa clic su OK, la volta successiva che il programma legge i dati (ad esempio, alla successiva esecuzione di una procedura statistica) viene creata una copia dei dati che consente di evitare un ulteriore lettura degli stessi. Se si fa clic su In cache ora, si ottiene la creazione immediata di

una copia dei dati, non necessaria nella maggior parte dei casi. In cache ora è utile soprattutto per due ragioni:

- Una sorgente dati è "bloccata" e non può essere aggiornata da nessuno fino al termine della sessione, all'apertura di un'altra sorgente dati o all'inserimento dei dati nella cache.
- Nel caso di sorgenti dati di grandi dimensioni, lo scorrimento del contenuto della Visualizzazione dati nell'Editor dei dati è molto più veloce se questi vengono inseriti nella cache.

Per eseguire una copia automatica dei dati

È possibile utilizzare il comando `SET` per creare automaticamente una copia dei dati dopo un numero specifico di modifiche nel file di dati attivo. Per impostazione predefinita, il file di dati attivo viene copiato automaticamente dopo 20 modifiche in tale file.

▶ Dai menu, scegliere:

File
  Nuovo
    Sintassi

▶ Nella finestra di sintassi, digitare SET CACHE n (dove *n* indica il numero di modifiche apportate al file di dati attivo prima di copiarlo nella cache).

▶ Dai menu della finestra di sintassi, scegliere:

Esegui
  Tutto

Nota: l'impostazione della cache non è valida in tutte le sessioni. A ogni avvio di una nuova sessione, il valore viene reimpostato al valore predefinito, ovvero 20.

Capitolo
4

Analisi distribuita

La modalità di analisi distribuita consente di utilizzare un computer che non sia quello locale (o desktop) per l'esecuzione di attività che prevedono un uso intensivo della memoria. Poiché i server remoti utilizzati per l'analisi distribuita sono in genere più potenti e veloci del computer locale, l'analisi distribuita può ridurre significativamente i tempi di elaborazione. L'analisi distribuita con un server remoto può rivelarsi vantaggiosa quando l'elaborazione prevede:

- L'utilizzo di file di dati di grandi dimensioni, in particolare dati letti da database.
- Attività che richiedono un uso intensivo della memoria. Tutte le attività che presuppongono tempi di elaborazione lunghi in modalità di analisi locale possono essere eseguite più velocemente in modalità di analisi distribuita.

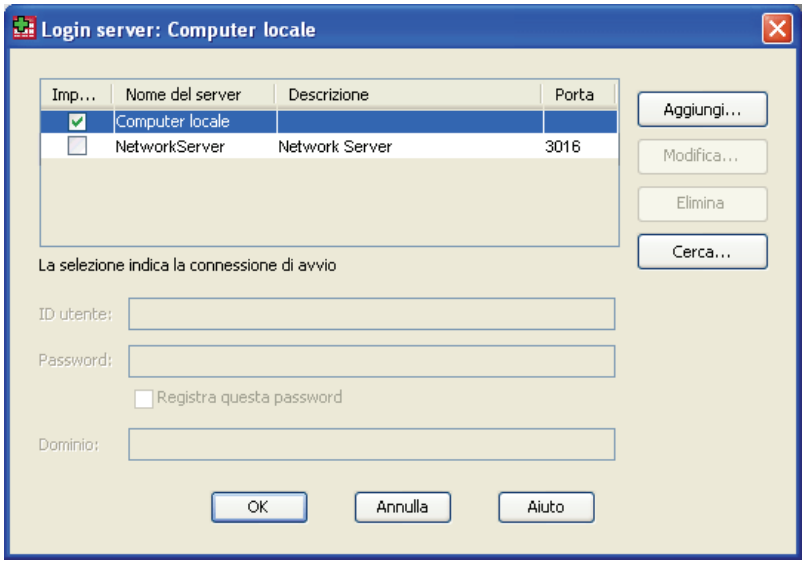
L'analisi distribuita è efficace solo nel caso di attività connesse ai dati, come lettura e trasformazione di dati, elaborazione di nuove variabili e calcolo di statistiche, mentre non lo è nel caso di attività connesse alla modifica dell'output, quali manipolazione di tabelle pivot o modifica di grafici.

Nota: l'analisi distribuita può essere utilizzata solo se si dispone di una versione locale del software e si ha accesso a una versione server con licenza dello stesso software installato su un server remoto.

Login al server

La finestra di dialogo Login server consente di selezionare il computer che elabora i comandi ed esegue le procedure, È possibile selezionare il computer locale o un server remoto.

Figura 4-1
Finestra di dialogo Login server



È possibile aggiungere, modificare o eliminare server remoti dall'elenco. I server remoti richiedono in genere un ID utente e una password, ma in certi casi anche il nome di dominio. Per informazioni relative a server disponibili, ID utente e password, nomi di dominio e altre informazioni per la connessione, rivolgersi all'amministratore di sistema.

È possibile selezionare un server predefinito e memorizzare ID utente, nome di dominio e password associati a qualsiasi server. All'avvio di una nuova sessione, viene eseguita automaticamente la connessione al server predefinito.

Se si è autorizzati a utilizzare Adaptor for Predictive Enterprise Services e sul sito è in esecuzione Predictive Enterprise Services 3.5 o versione successiva, è possibile fare clic su Cerca... per visualizzare l'elenco dei server disponibili nella rete. Se non si è collegati a un Predictive Enterprise Repository, per la visualizzazione dell'elenco dei server verrà richiesto di immettere le informazioni sulla connessione.

Aggiunta e modifica delle impostazioni di login al server

Utilizzare la finestra di dialogo Impostazioni login al server per aggiungere o modificare le informazioni per la connessione ai server remoti da utilizzare in modalità di analisi distribuita.

Figura 4-2
Finestra di dialogo Impostazioni login al server

Per un elenco dei server disponibili, dei numeri di porta per i server e altre informazioni per la connessione, rivolgersi all'amministratore del sistema. Non utilizzare Secure Sockets Layer a meno che non venga richiesto specificatamente dall'amministratore.

Nome server. Il "nome" di un server può essere un nome alfanumerico assegnato a un computer (ad esempio, NetworkServer) o un indirizzo IP univoco assegnato a un computer (ad esempio, 202.123.456.78).

Numero porta. Il numero di porta corrisponde alla porta che il software del server utilizza per le comunicazioni.

Descrizione. È possibile immettere una descrizione facoltativa da visualizzare nell'elenco dei server.

Connetti con Secure Socket Layer. Secure Socket Layer (SSL) consente di crittografare le richieste di analisi distribuita durante l'invio al server remoto. Prima di utilizzare SSL, contattare l'amministratore di sistema. Per attivare questa opzione, è necessario configurare SSL sul computer desktop e sul server.

Per selezionare, cambiare o aggiungere server

- Dai menu, scegliere:

  File
  Cambia server...

Per selezionare un server predefinito:

- Nell'elenco dei server selezionare la casella accanto al server che si desidera utilizzare.
- Immettere l'ID utente, il nome di dominio e la password forniti dall'amministratore.

  Nota: All'avvio di una nuova sessione, viene eseguita automaticamente la connessione al server predefinito.

Per cambiare server:

- Selezionare il server dall'elenco.
- Se necessario, immettere l'ID utente, il nome di dominio e la password.

Nota: Se durante una sessione si cambia server, tutte le finestre aperte vengono chiuse. Prima dell'effettiva chiusura delle finestre, viene chiesto se si desidera salvare le modifiche.

Per aggiungere un server:

- Rivolgersi all'amministratore per le informazioni di connessione al server.
- Fare clic su Aggiungi per aprire la finestra di dialogo Impostazioni login al server.
- Immettere le informazioni di connessione e le impostazioni opzionali e fare clic su OK.

Per modificare un server:

- Rivolgersi all'amministratore per le informazioni di connessione al server aggiornate.
- Fare clic su Modifica per aprire la finestra di dialogo Impostazioni login al server.
- Apportare le modifiche e fare clic su OK.

Per cercare i server disponibili:

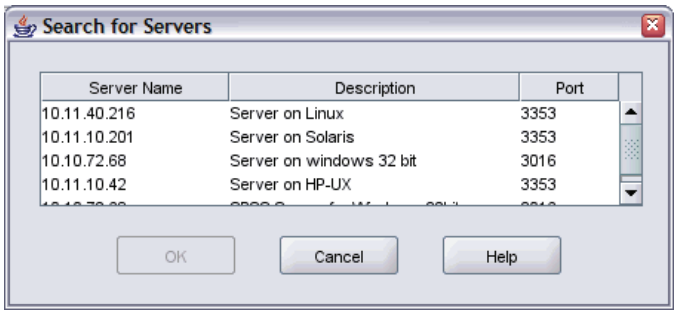
Nota: La funzione di ricerca dei server disponibili è abilitata solo se si è autorizzati ad utilizzare Adaptor for Predictive Enterprise Services e sul sito è in esecuzione Predictive Enterprise Services 3.5 o versione successiva.

- Fare clic su Cerca... per aprire la finestra di dialogo Cerca i server. Se non si è collegati a un Predictive Enterprise Repository, verrà richiesto di immettere le informazioni sulla connessione.
- Selezionare uno o più server disponibili e fare clic su OK. I server vengono visualizzati nella finestra di dialogo Login server.
- Per connettersi a uno dei server, seguire le istruzioni riportate in "Per cambiare server".

Ricerca dei server disponibili

Utilizzare la finestra di dialogo Cerca i server per selezionare uno o più server disponibili nella rete. La finestra di dialogo viene visualizzata quando si fa clic su Cerca... nella finestra di dialogo Login server.

Figura 4-3
Finestra di dialogo Cerca i server



Selezionare uno o più server e fare clic su OK per aggiungerli alla finestra di dialogo Login server. Sebbene sia possibile aggiungere manualmente i server nella finestra di dialogo Login server, la ricerca dei server disponibili consente di connettersi ai server senza che sia necessario conoscerne il nome e il numero di porta. Queste informazioni vengono fornite automaticamente. Tuttavia, è comunque possibile correggere le informazioni sull'accesso, quali il nome utente, il dominio e la password.

Apertura di file di dati da un server remoto

Nell'analisi distribuita la finestra di dialogo Apri file remoto sostituisce la finestra di dialogo Apri file standard.

- Il contenuto dell'elenco di file, cartelle e unità disponibili varia a seconda di quanto risulta disponibile sul server remoto. Il nome del server corrente è indicato nella parte superiore della finestra di dialogo.
- In modalità di analisi distribuita, non sarà possibile accedere ai file di dati del computer locale se non si imposta la condivisione dell'unità e delle cartelle contenenti quei file. Per informazioni sulla condivisione delle cartelle del computer locale con la rete di server, consultare la documentazione del proprio sistema operativo.
- Se il server esegue un sistema operativo diverso (ad esempio, sul proprio computer è installato Windows e il server esegue UNIX), è probabile che non si possa accedere ai file di dati locali in modalità di analisi distribuita, anche se si trovano in cartelle condivise.

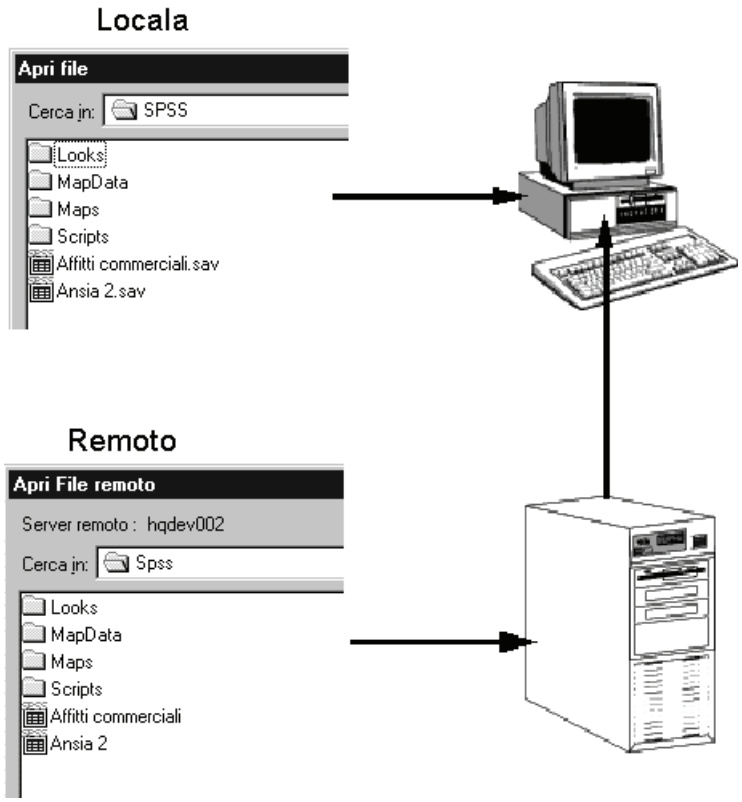
Accesso ai file in modalità di analisi locale e distribuita

La modalità di visualizzazione dei file di dati, delle cartelle (directory) e delle unità nel computer locale e nella rete dipende dal computer in cui vengono eseguite l'elaborazione dei comandi e le procedure, che non è necessariamente il computer locale.

Modalità di analisi locale. Quando si utilizza il computer locale come "server", la modalità di visualizzazione di file di dati, cartelle e unità nella finestra di dialogo di accesso ai file (per l'apertura dei file di dati) è simile a quella di altre applicazioni o di Esplora risorse di Windows. È possibile visualizzare tutti i file di dati e le cartelle del computer, nonché i file e le cartelle delle unità di rete installate e in genere visualizzate.

Modalità di analisi distribuita. Quando si utilizza un altro computer come "server remoto" per l'esecuzione di comandi e procedure, i file di dati, le cartelle e le unità vengono visualizzati secondo la prospettiva del server remoto. Sebbene i nomi delle cartelle e delle unità come *Programmi* e *C* possano sembrare familiari, questi *non* si riferiscono alle cartelle e alle unità presenti sul computer locale; ma alle cartelle e alle unità del server remoto.

Figura 4-4
Visualizzazioni locali e remote



In modalità di analisi distribuita, non sarà possibile accedere ai file di dati del computer locale se non si imposta la condivisione dell'unità e delle cartelle contenenti quei file. Se il server esegue un sistema operativo diverso (ad esempio, sul proprio computer è installato Windows e il server esegue UNIX), è probabile che non si possa accedere ai file di dati locali in modalità di analisi distribuita, anche se si trovano in cartelle condivise.

La modalità di analisi distribuita non equivale all'accesso ai file di dati che risiedono su un altro computer della rete. È possibile accedere ai file di dati presenti su altre periferiche di rete sia in modalità di analisi locale sia in modalità di analisi distribuita. In modalità locale è possibile accedere ad altre periferiche di rete dal computer locale. In modalità distribuita vi si accede dal server remoto.

Se non si è certi della modalità di analisi in uso, leggere la barra del titolo della finestra di dialogo per l'accesso ai file di dati. Se la barra del titolo contiene la parola *Remoto* (come in Apri file remoto) oppure se nella parte superiore della finestra di dialogo compare la dicitura Server remoto: [nome server], si sta utilizzando la modalità di analisi distribuita.

Nota: Ciò vale solo per le finestre di dialogo utilizzate per l'accesso ai file di dati (ad esempio, Apri dati, Salva dati, Apri database e Applica dizionario dati). Per tutti gli altri tipi (ad esempio, file Viewer, file di sintassi e file script), si utilizza sempre la visualizzazione locale.

Procedure disponibili in modalità di analisi distribuita

In modalità di analisi distribuita, sono disponibili solo le procedure installate sia nella versione locale che nella versione del server remoto.

Se sul computer locale sono stati installati componenti non disponibili sul server remoto e si passa dal computer locale al server remoto, le procedure in questione verranno rimosse dai menu. Di conseguenza, le sintassi di comando corrispondenti genereranno errori. Se si torna alla modalità locale, tutte le procedure interessate vengono ripristinate.

Specifiche di percorsi assoluti e relativi

In modalità di analisi distribuita, le specifiche di percorso relativo per i file di dati e i file di sintassi dei comandi sono relative al server corrente e non al computer locale. Ciò significa che un percorso relativo, quale */mydocs/mydata.sav*, non punta a una directory e a un file presenti sull'unità locale, ma a una directory e a un file sul disco rigido del server remoto.

Specifiche di percorso UNC di Windows

Se si sta eseguendo una versione del server Windows, è possibile utilizzare le specifiche UNC (Universal Naming Convention, convenzione universale di denominazione) quando si accede ai file di dati e di sintassi con la sintassi dei comandi. La forma generica di una specifica UNC è la seguente:

```
\\nomeserver\condivisione\percorso\nomefile
```

- *Nomeserver* è il nome del computer in cui si trova il file di dati.
- *Condivisione* è la cartella (directory) su quel computer designata come cartella condivisa.
- *Percorso* è il percorso delle eventuali sottocartelle (sottodirectory) della cartella condivisa.
- *Nomefile* è il nome del file di dati.

Esempio:

```
GET FILE = '\\hqdev01\public\luglio\vendite.sav'.
```

Se al computer non è stato assegnato un nome, è possibile utilizzare un indirizzo IP, ad esempio:

```
GET FILE='\\204.125.125.53\public\luglio\vendite.sav'.
```

Anche con le specifiche di percorso UNC, è possibile accedere ai file di dati e di sintassi presenti unicamente nelle periferiche e nelle cartelle condivise. Quando si utilizza la modalità di analisi distribuita, ciò vale anche per i file di dati e di sintassi presenti sul computer locale.

Specifiche di percorsi assoluti UNIX

Per le versioni di server UNIX, non esiste un equivalente del percorso UNC e tutti i percorsi alle directory devono essere percorsi assoluti che partono dalla directory principale del server. I percorsi relativi non sono consentiti. Ad esempio, se il file di dati si trova in */bin/data* e anche la directory corrente è */bin/data*, `GET FILE='sales.sav'` non è valido in quanto è necessario specificare tutto il percorso come segue:

```
GET FILE='/bin/sales.sav'.
```

```
INSERT FILE='/bin/salesjob.sps'.
```

Capitolo
5

Editor dei dati

L'Editor dei dati è uno strumento utile per la creazione e la modifica di file di dati, analogo a quello dei fogli di calcolo. La finestra dell'Editor dei dati viene aperta automaticamente all'apertura di ogni sessione di SPSS.

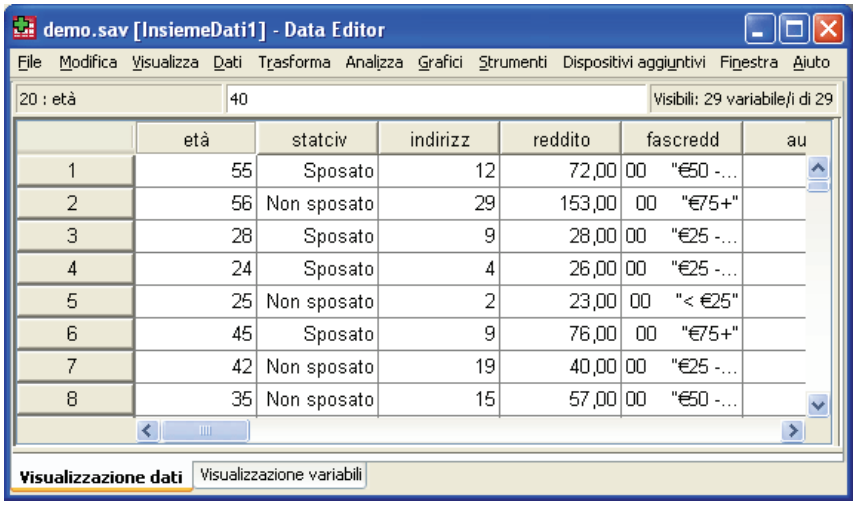
L'Editor dei dati offre due visualizzazioni dei dati:

- **Visualizzazione dati.** Visualizza i valori dei dati effettivi o le etichette dei valori definite.
- **Visualizzazione variabili.** Visualizza le informazioni di definizione delle variabili, incluse le etichette delle variabili e dei valori definite, il tipo di dati (ad esempio, stringa, data o numerico), il livello di misurazione (nominale, ordinale o scala) e i valori mancanti definiti dall'utente.

In entrambe le visualizzazioni, è possibile aggiungere, modificare ed eliminare le informazioni contenute nel file di dati.

Visualizzazione dati

Figura 5-1
Visualizzazione dati


demo.sav [InsiemeDati1] - Data Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntivi Finestra Aiuto
20 : età
40
Visibili: 29 variabile/i di 29
età statciv indirizz reddito fascredd au
1 55 Sposato 12 72,00 00 "€50 -...
2 56 Non sposato 29 153,00 00 "€75+"
3 28 Sposato 9 28,00 00 "€25 -...
4 24 Sposato 4 26,00 00 "€25 -...
5 25 Non sposato 2 23,00 00 "< €25"
6 45 Sposato 9 76,00 00 "€75+"
7 42 Non sposato 19 40,00 00 "€25 -...
8 35 Non sposato 15 57,00 00 "€50 -...
Visualizzazione dati
Visualizzazione variabili

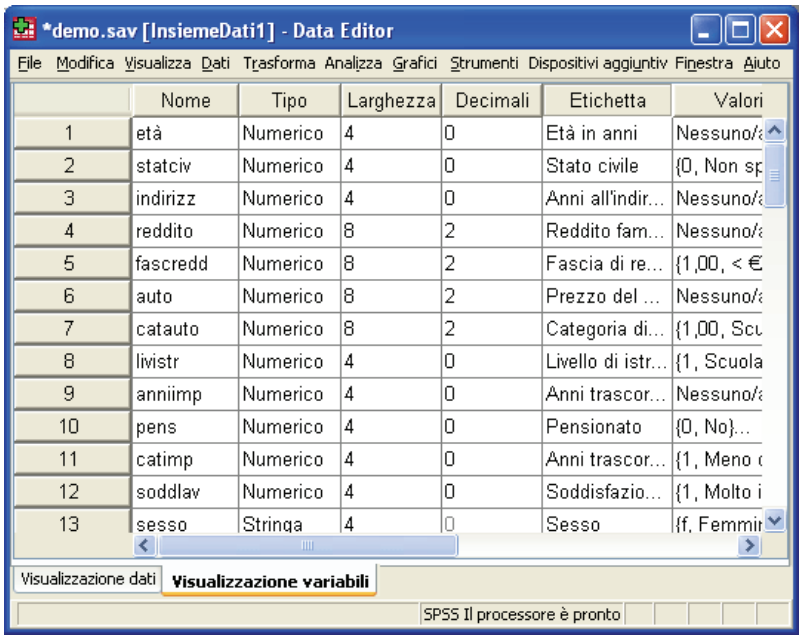
Molte funzioni di Visualizzazione dati sono analoghe a quelle disponibili nelle applicazioni per fogli di calcolo. Esistono, tuttavia, diverse importanti distinzioni:

- Le righe sono i casi. Ciascuna riga rappresenta un caso o un'osservazione. Ogni persona che risponde a un questionario, ad esempio, rappresenta un caso.

- Le colonne sono le variabili. Ciascuna colonna rappresenta una variabile o una caratteristica da misurare. Ogni voce del questionario, ad esempio, rappresenta una variabile.
- Le celle contengono valori. Ogni cella contiene un singolo valore di una variabile relativa a un caso. La cella è il punto in cui il caso e la variabile si intersecano. Le celle contengono solo valori. A differenza dei programmi per fogli di calcolo, le celle dell'Editor dei dati non possono contenere formule.
- Il file di dati è rettangolare. Le dimensioni del file di dati vengono determinate dal numero di casi e di variabili. È possibile inserire dati in qualsiasi cella. Se si inseriscono dati in una cella che non rientra nei limiti del file di dati definito, il rettangolo di dati verrà esteso in modo da includere le righe e/o le colonne che si trovano tra la cella e i limiti del file. Entro i limiti del file di dati non sono presenti celle vuote. Per le variabili numeriche, le celle vuote vengono convertite al valore mancante di sistema. Per le variabili stringa, le celle vuote vengono considerate valori validi.

Visualizzazione variabili

Figura 5-2
Visualizzazione variabili


*demo.sav [InsiemeDati1] - Data Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntiv Finestra Aiuto
Nome Tipo Larghezza Decimali Etichetta Valori
1 età Numerico 4 0 Età in anni Nessuno/
2 statciv Numerico 4 0 Stato civile {0, Non sp
3 indirizz Numerico 4 0 Anni all'indir... Nessuno/
4 reddito Numerico 8 2 Reddito fam... Nessuno/
5 fascredd Numerico 8 2 Fascia di re... {1,00, < €
6 auto Numerico 8 2 Prezzo del ... Nessuno/
7 catauto Numerico 8 2 Categoria di... {1,00, Scu
8 livistr Numerico 4 0 Livello di istr... {1, Scuola
9 anniimp Numerico 4 0 Anni trascor... Nessuno/
10 pens Numerico 4 0 Pensionato {0, No}...
11 catimp Numerico 4 0 Anni trascor... {1, Meno
12 soddlav Numerico 4 0 Soddisfazio... {1, Molto i
13 sesso Stringa 4 0 Sesso {f, Femmi
Visualizzazione dati
Visualizzazione variabili
SPSS Il processore è pronto

Visualizzazione variabili contiene le descrizioni degli attributi di ciascuna variabile del file di dati. In Visualizzazione variabili:

- Le righe sono le variabili.
- Le colonne sono gli attributi delle variabili.

È possibile aggiungere o eliminare variabili e modificare attributi delle variabili, tra cui:

- Nome di variabile
- Tipo di dati

- Numero di cifre o di caratteri
- Numero di posizioni decimali
- Etichette descrittive delle variabili e dei valori
- Valori mancanti definiti dall'utente
- Larghezza di colonna
- Livello di misurazione

Tutti questi attributi vengono salvati con il file di dati.

Oltre alla definizione delle proprietà delle variabili disponibile in Visualizzazione variabili, esistono due ulteriori metodi per definire queste proprietà:

- La Copia proprietà dei dati guidata permette di usare un file di dati di SPSS Statistics esterno o un altro file di dati disponibile nella sessione corrente come modello per la definizione delle proprietà dei file e delle variabili nel file di dati attivo. È possibile anche usare le variabili nell'insieme di dati attivo come modelli per altre variabili dello stesso insieme di dati. Copia proprietà dei dati è disponibile nel menu Dati della finestra dell'Editor dei dati.
- Definisci proprietà variabili (disponibile anche nel menu Dati della finestra dell'Editor dei dati) esamina i dati ed elenca tutti i valori univoci dei dati per ogni variabile selezionata, inoltre identifica i valori senza etichetta e fornisce una funzione per l'aggiunta automatica dell'etichetta. Questo risulta particolarmente utile per le variabili categoriali che utilizzano codici numerici per rappresentare le categorie, ad esempio 0 = *Maschio*, 1 = *Femmina*.

Per visualizzare e definire gli attributi delle variabili

- ▶ Attivare la finestra dell'Editor dei dati.
- ▶ Fare doppio clic su un nome di variabile nella parte superiore della colonna in Visualizzazione dati oppure fare clic sulla scheda Visualizzazione variabili.
- ▶ Per definire una nuova variabile, immettere un nome di variabile in una qualsiasi riga vuota.
- ▶ Selezionare gli attributi da definire o modificare.

Nomi delle variabili

Per i nomi delle variabili sono valide le seguenti regole:

- Ogni variabile deve essere univoca. La duplicazione non è consentita.
- I nomi di variabile possono essere lunghi fino a 64 byte e il primo carattere deve essere una lettera o uno dei seguenti caratteri: @, # o $. I caratteri successivi possono essere una qualsiasi combinazione di lettere, numeri, caratteri non appartenenti alla punteggiatura e un punto (.). In modalità di pagina di codice, sessantaquattro byte in genere corrispondono a 64 caratteri nelle lingue a byte singolo (ad esempio, inglese, francese, tedesco, spagnolo, italiano, ebraico, russo, greco, arabo e tailandese) e a 32 caratteri nelle lingue a doppio byte (giapponese, cinese e coreano). Molti caratteri stringa che utilizzano un byte in modalità di pagina di codice, utilizzano due o più byte in modalità Unicode. Ad esempio, é è un byte in

formato di codice di pagina ma due byte in formato Unicode; quindi *résumé* corrisponde a sei byte in modalità di pagina di codice e otto byte in modalità Unicode.

Nota: le lettere includono qualsiasi carattere non appartenente alla punteggiatura utilizzato nella scrittura di parole comuni nelle lingue supportate nel set di caratteri della piattaforma.

- I nomi di variabile non possono contenere spazi.
- Il carattere # in posizione iniziale nel nome di una variabile definisce una variabile vuota. Le variabili vuote possono essere create solo tramite la sintassi dei comandi. Non è consentito specificare # in posizione iniziale per una variabile nelle finestre di dialogo di creazione di nuove variabili.
- Il simbolo $ in posizione iniziale indica che si tratta di una variabile di sistema. Il simbolo $ non è ammesso come carattere iniziale di una variabile definita dall'utente.
- Il punto, la sottolineatura e i caratteri $, # e @ possono essere usati all'interno dei nomi di variabile. Ad esempio, *A._$@#1* è un nome di variabile valido.
- I nomi di variabile che terminano con un punto devono essere evitati, poiché il punto può essere interpretato come la fine del comando. Le variabili con punto in posizione finale possono essere create solo tramite nella sintassi dei comandi. Non è consentito creare variabili con punto in posizione finale nelle finestre di dialogo di creazione di nuove variabili.
- I nomi di variabile che terminano con sottolineatura devono essere evitati, poiché tali nomi possono creare un conflitto con nomi di variabili automaticamente create da comandi e procedure.
- Le parole chiave standard non possono essere utilizzate come nomi di variabili. Le parole chiave standard sono ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO e WITH.
- È possibile definire i nomi delle variabili con un misto di caratteri maiuscoli e minuscoli e la distinzione viene mantenuta per motivi di visualizzazione.
- Quando è necessario impostare il ritorno a capo su più righe dell'output per i nomi di variabili lunghi, le righe vengono interrotte in corrispondenza dei caratteri di sottolineatura, dei punti e delle occorrenze di contenuto in cui i caratteri minuscoli sono stati modificati in maiuscolo.

Livello di misurazione delle variabili

È possibile specificare il livello di misurazione in forma di scala (dati numerici su una scala di intervallo o di rapporto), come ordinale o nominale. I dati nominali e ordinali possono essere rappresentati da stringhe (alfanumeriche) o essere numerici.

- **Nominale.** Una variabile può essere considerata nominale quando i relativi valori rappresentano categorie prive di ordinamento intrinseco, per esempio l'ufficio di una società, la regione, il codice postale e la religione.

- **Ordinale.** Una variabile può essere considerata ordinale quando i relativi valori rappresentano categorie con qualche ordinamento intrinseco, per esempio i gradi di soddisfazione per un servizio, da molto insoddisfatto a molto soddisfatto, i punteggi di atteggiamento corrispondenti a gradi di soddisfazione o fiducia e i punteggi di preferenza.
- **Scala.** Una variabile può essere considerata di scala quando i relativi valori rappresentano categorie ordinate con una metrica significativa, tale che i confronti fra le distanze dei relativi valori siano appropriati. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

Nota: per le variabili stringa ordinali si suppone che l'ordine alfabetico dei valori stringa rifletta l'esatto ordine delle categorie. Ad esempio, per una variabile stringa con i valori *basso*, *medio*, *alto*, l'ordine delle categorie viene interpretato come *alto*, *basso*, *medio*, ma questo non è l'ordine corretto. In generale, per rappresentare i dati ordinali, è più sicuro utilizzare i codici numerici.

Alle nuove variabili numeriche create durante una sessione viene assegnato il livello di misurazione Scala. Per i dati letti da formati di file esterni e dai file di dati di SPSS Statistics che sono stati creati con versioni precedenti alla 8.0, il livello di misurazione predefinito assegnato è basato sulle seguenti regole:

- Le variabili numeriche con meno di 24 valori univoci e le variabili stringa sono impostate su Nominale.
- Le variabili numeriche con 24 o più valori univoci sono impostate su Scala.

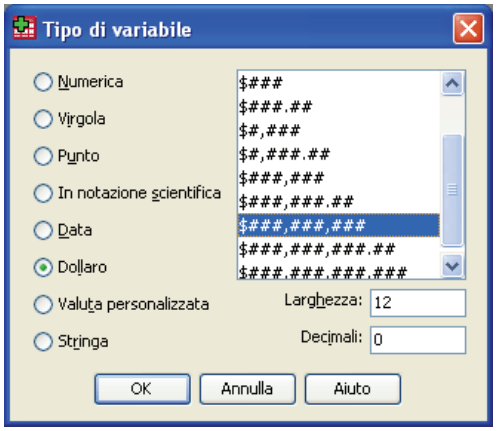
Il valore di riferimento scala/nominale per le variabili scala/nominale può essere modificato nella finestra di dialogo Opzioni. Per ulteriori informazioni, vedere Opzioni: Dati in Capitolo 48 a pag. 525.

La finestra di dialogo Definisci proprietà variabili nel menu Dati permette di assegnare più agevolmente il livello di misurazione corretto. Per ulteriori informazioni, vedere Assegnazione del livello di misurazione in Capitolo 7 a pag. 111.

Tipo di variabile

Tipo di variabile consente di specificare il tipo di dati per ciascuna variabile. Per impostazione predefinita, si presume che tutte le nuove variabili siano numeriche. È possibile utilizzare Tipo di variabile per modificare il tipo di dati. Il contenuto della finestra di dialogo Tipo di variabile varia in base al tipo di dati selezionato. Per alcuni tipi di dati, sono disponibili caselle di testo per immettere la larghezza e il numero di decimali, mentre per altri è sufficiente selezionare un formato da un elenco di esempi.

Figura 5-3
Finestra di dialogo Tipo di variabile



I tipi di dati disponibili sono:

Numerica. Una variabile che contiene solo dati numerici. visualizzati nel formato numerico standard. L'Editor dei dati accetta valori numerici nel formato standard oppure in notazione scientifica.

Virgola. Una variabile numerica i cui valori sono delimitati con una virgola ogni tre posizioni e con il punto come delimitatore dei decimali. Per questo tipo di variabili l'Editor dei dati accetta valori numerici con o senza virgola oppure valori in notazione scientifica. I valori non possono contenere virgole a destra dell'indicatore decimale.

Punto. Una variabile numerica i cui valori sono delimitati con un punto ogni tre posizioni e con la virgola come delimitatore dei decimali. L'Editor dei dati accetta valori numerici per variabili punto con o senza punti oppure li accetta in notazione scientifica. I valori non possono contenere punti a destra dell'indicatore decimale.

In notazione scientifica. Una variabile numerica i cui valori sono visualizzati con una E e un esponente alla potenza 10. L'Editor dei dati accetta valori numerici per tali variabili con o senza esponente. L'esponente può essere preceduto da E o D con un simbolo opzionale o dal solo simbolo, ad esempio 123, 1.23E2, 1.23D2, 1.23E+2 o anche 1.23+2.

Data. Una variabile numerica i cui valori sono visualizzati in uno dei formati calendario-data o orologio-tempo. Selezionare un formato dall'elenco. I delimitatori di data accettati sono barre, trattini, punti, virgole o spazi vuoti. L'intervallo di secoli per i valori di anni a due cifre viene determinato mediante le impostazioni delle opzioni (menu Modifica, Opzioni, schedaDati).

Dollaro. Una variabile numerica visualizzata con un simbolo dollaro iniziale ($), con virgole di delimitazione per ogni tre spazi e un punto come delimitatore decimale. È possibile immettere i valori dati con o senza il simbolo dollaro iniziale.

Valuta personalizzata. Una variabile numerica i cui valori sono visualizzati in uno dei formati di valuta personalizzati definiti nella scheda Valuta della finestra di dialogo Opzioni. I caratteri di valuta personalizzati definiti non possono essere utilizzati nell'inserimento dei dati ma vengono visualizzati nell'Editor dei dati.

Stringa. I valori di una variabile stringa non sono numerici e non vengono quindi utilizzati per i calcoli. Queste variabili possono contenere qualsiasi carattere e non devono superare la lunghezza definita. È attiva la distinzione tra maiuscole e minuscole. Questo tipo di variabili è chiamato anche variabili alfanumeriche.

Per definire il tipo di variabile

- ▶ Fare clic sul pulsante della cella *Tipo* della variabile che si intende definire.
- ▶ Selezionare il tipo di dati desiderato nella finestra di dialogo Tipo di variabile.
- ▶ Fare clic su OK.

Formati di input e di visualizzazione

A seconda del formato, i valori in Visualizzazione dati possono differire dal valore effettivo inserito e memorizzato internamente. Di seguito vengono riportate alcune indicazioni generali:

- Per i formati numerico, virgola e punto, è possibile specificare valori con qualsiasi numero di cifre decimali (fino a 16). Il valore verrà memorizzato per intero. Nella Visualizzazione dati verrà visualizzato solo il numero specificato di cifre decimali. I valori con cifre decimali in eccesso verranno arrotondati. In qualsiasi calcolo verrà tuttavia utilizzato il valore completo.
- Per le variabili stringa, a tutti i valori viene aggiunto il numero di spazi necessario per raggiungere la larghezza massima. Per una variabile stringa con una larghezza massima pari a tre, il valore *No* viene salvato internamente come `'No '` e non è equivalente a `' No'`.
- Per i formati della data, è possibile utilizzare barre, trattini, spazi, virgole o punti come delimitatori tra i valori relativi a giorno, mese e anno, nonché aggiungere numeri, abbreviazioni di tre lettere o nomi completi per i valori relativi ai mesi. Le date espresse nel formato generico *gg-mmm-aa* vengono visualizzate con il trattino come delimitatore e l'abbreviazione di tre lettere del mese. Le date espresse nel formato generico *gg/mm/aa* e *mm/gg/aa* vengono visualizzate con la barra come delimitatore e i numeri corrispondenti ai mesi. Le date vengono memorizzate internamente come numero di secondi trascorsi dal 14 ottobre 1582. L'intervallo di secoli per gli anni a due cifre viene determinato mediante le impostazioni delle opzioni (menu Modifica, Opzioni, scheda Dati).
- Per i formati dell'orario, è possibile utilizzare i due punti, il punto o lo spazio come delimitatore tra ore, minuti e secondi. Gli orari vengono visualizzati con i due punti come delimitatore. Internamente le date vengono memorizzate come il numero di secondi che rappresentano un intervallo di tempo. Ad esempio, 10:00:00 è memorizzato internamente come 36000, ossia 60 (secondi/minuto) x 60 (minuti/ora) x 10 (ore).

Etichette delle variabili

È possibile assegnare etichette variabili descrittive della lunghezza massima di 256 caratteri (o 128 caratteri nelle lingue a due byte). Le etichette variabili possono contenere spazi e caratteri riservati non consentiti nei nomi delle variabili.

Per specificare le etichette di variabile

- Attivare la finestra dell'Editor dei dati.
- Fare doppio clic su un nome di variabile nella parte superiore della colonna in Visualizzazione dati oppure fare clic sulla scheda Visualizzazione variabili.
- Immettere l'etichetta di variabile descrittiva nella cella *Etichetta*.

Etichette dei valori

È possibile assegnare etichette descrittive a ciascun valore di una variabile. Questo processo risulta particolarmente utile se il file di dati utilizza codici numerici per rappresentare categorie non numeriche, ad esempio i codici 1 e 2 per *maschio* e *femmina*.

- Le etichette dei valori possono essere formate al massimo da 120 byte.

Figura 5-4
Finestra di dialogo Etichette dei valori

Per specificare le etichette dei valori

- Fare clic sul pulsante della cella *Valori* della variabile che si intende definire.
- Per ciascun valore, specificare il valore e l'etichetta corrispondente.
- Fare clic su Aggiungi per inserire l'etichetta del valore.
- Fare clic su OK.

Inserimento di interruzioni di riga nelle etichette

Per le etichette di variabili e dei valori, viene impostato il ritorno a capo automatico su più righe nelle tabelle pivot e nei grafici se la larghezza della cella o dell'area non consente di visualizzare l'intera etichetta su un'unica linea. Per spostare il ritorno a capo in un punto diverso dell'etichetta, è possibile modificare i risultati e inserire manualmente le interruzioni di riga. È inoltre possibile

creare etichette di variabili e dei valori in cui il ritorno a capo si verificherà *sempre* in determinati punti con conseguente visualizzazione su più righe:

- Per le etichette delle variabili, selezionare la cella *Etichetta* per la variabile nella Visualizzazione variabili dell'Editor dei dati.
- Per le etichette dei valori, selezionare la cella *Valori* per la variabile in Visualizzazione variabili dell'Editor dei dati, fare clic sul pulsante visualizzato nella cella e selezionare l'etichetta che si desidera modificare nella finestra di dialogo Etichette dei valori.
- Digitare \n all'interno dell'etichetta nel punto in cui inserire l'interruzione di riga.

  Il simbolo \n non viene visualizzato nelle tabelle pivot o nei grafici, ma viene considerato un carattere indicante l'interruzione di riga.

Valori mancanti

Definisci valori mancanti consente di determinare i valori dei dati specificati come **valori mancanti definiti dall'utente**. Può essere necessario, ad esempio, distinguere tra dati che mancano perché la persona ha rifiutato di rispondere e dati che mancano perché la domanda non riguardava la persona in questione. I dati specificati come valori mancanti definiti dall'utente vengono evidenziati in modo da poter essere elaborati diversamente ed esclusi dalla maggior parte dei calcoli.

Figura 5-5
Finestra di dialogo Valori mancanti

- È possibile inserire al massimo tre valori mancanti discreti (singoli), un intervallo di valori mancanti oppure un intervallo più un valore discreto.
- È possibile specificare intervalli solo per le variabili numeriche.
- Tutti i valori delle stringhe, inclusi i valori nulli o vuoti, sono considerati validi a meno che non siano esplicitamente definiti mancanti.
- I valori mancanti delle variabili stringa non possono superare gli otto byte. Non è previsto alcun limite per la larghezza definita della variabile di tipo stringa, ma i valori mancanti definiti non possono superare gli otto byte.
- Per definire come valori mancanti i valori nulli o vuoti di una variabile stringa, immettere uno spazio singolo in uno dei campi per Valori mancanti discreti.

Per definire i valori mancanti

- ▶ Fare clic sul pulsante della cella *Mancante* per la variabile che si intende definire.
- ▶ Inserire i valori o l'intervallo di valori che rappresentano i dati mancanti.

Larghezza colonna

È possibile specificare un numero di caratteri per la larghezza della colonna. Per modificare la larghezza delle colonne in Visualizzazione dati, fare clic e trascinare i bordi delle colonne.

- La larghezza della colonna per i caratteri proporzionali è basata sulla larghezza media dei caratteri. A seconda dei caratteri utilizzati nel valore, possono essere visualizzati più o meno caratteri nella larghezza specificata.
- La larghezza della colonna influisce solo sulla visualizzazione dei valori nell'Editor dei dati. La modifica della larghezza delle colonne non comporta la modifica della larghezza definita di una variabile.

Allineamento

Allineamento consente di controllare la visualizzazione dei valori dei dati e/o delle etichette dei valori in Visualizzazione dati. L'allineamento predefinito è a destra per le variabili numeriche e a sinistra per le variabili stringa. L'impostazione influisce esclusivamente su Visualizzazione dati.

Applicazione degli attributi di definizione di una variabile ad altre variabili

Dopo avere specificato gli attributi di definizione di una variabile, è possibile copiare uno o più attributi e applicarli a una o più variabili.

Per applicare gli attributi di definizione di una variabile, è possibile utilizzare le operazioni di copia e incolla di base. È possibile:

- Copiare un singolo attributo (ad esempio, le etichette dei valori) e incollarlo nella cella del medesimo attributo di una o più variabili.
- Copiare tutti gli attributi di una variabile e incollarli in una o più variabili diverse.
- Creare più variabili nuove con tutti gli attributi di una variabile copiata.

Applicazione degli attributi di definizione di una variabile ad altre variabili

Per applicare dei singoli attributi di una determinata variabile

- ▶ In Visualizzazione variabili, selezionare la cella dell'attributo da applicare alle altre variabili.
- ▶ Dai menu, scegliere:

  Modifica
  Copia

- ▶ Selezionare la cella a cui applicare l'attributo. È possibile selezionare più variabili di destinazione.

- Dai menu, scegliere:
  Modifica
  Incolla

  Se si incolla l'attributo in righe vuote, vengono create nuove variabili con i valori predefiniti per tutti gli attributi fatta eccezione per l'attributo selezionato.

Per applicare tutti gli attributi di una determinata variabile

- In Visualizzazione variabili, selezionare il numero di riga della variabile con gli attributi da utilizzare. Verrà evidenziata l'intera riga.
- Dai menu, scegliere:
  Modifica
  Copia
- Selezionare i numeri di riga delle variabili a cui applicare gli attributi. È possibile selezionare più variabili di destinazione.
- Dai menu, scegliere:
  Modifica
  Incolla

Creazione di più variabili nuove con gli stessi attributi

- In Visualizzazione variabili, fare clic sul numero di riga della variabile con gli attributi da utilizzare per la nuova variabile. Verrà evidenziata l'intera riga.
- Dai menu, scegliere:
  Modifica
  Copia
- Fare clic sul numero di riga vuota sotto l'ultima variabile definita nel file di dati.
- Dai menu, scegliere:
  Modifica
  Incolla variabili
- Nella finestra di dialogo Incolla variabili, immettere il numero di variabili da creare.
- Inserire un prefisso e un numero iniziali per le nuove variabili.
- Fare clic su OK.

  I nomi delle nuove variabili saranno composti dal prefisso specificato a cui viene aggiunto un numero sequenziale che inizia con il numero specificato.

Attributi variabili personalizzati

Oltre agli attributi standard delle variabili (ad esempio, etichette di valore, valori mancanti, livello di misurazione), è possibile creare degli attributi di variabile personalizzati. Analogamente agli attributi standard delle variabili, questi attributi personalizzati vengono salvati in file di dati in formato SPSS Statistics. Pertanto, è possibile creare un attributo di variabile che identifichi il tipo

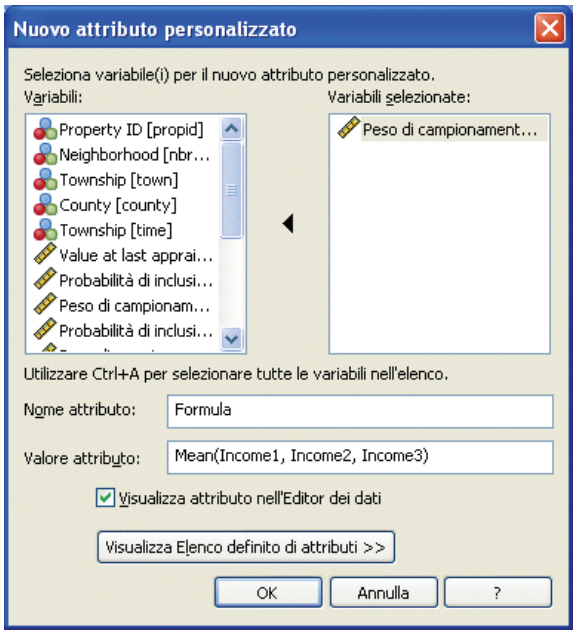
di risposta alle domande di un'indagine (ad esempio a scelta singola, multipla o con riempimento degli spazi vuoti) oppure le formule utilizzate per le variabili calcolate.

Creazione di attributi di variabile personalizzati

Per creare nuovi attributi personalizzati:

- Dai menu della Visualizzazione variabili scegliere:
  Dati
  Nuovo attributo personalizzato...
- Trascinare le variabili a cui assegnare il nuovo attributo nell'elenco Variabili selezionate.
- Digitare un nome per l'attributo. I nomi di attributo devono rispettare le stesse regole applicate ai nomi di variabile. Per ulteriori informazioni, vedere Nomi delle variabili a pag. 79.
- Digitare un valore facoltativo per l'attributo. Se la selezione copre più variabili, il valore viene assegnato a tutte quelle selezionate. È possibile lasciare questo campo vuoto e digitare un valore per ciascuna variabile in Visualizzazione variabili.

Figura 5-6
Finestra di dialogo Nuovo attributo personalizzato



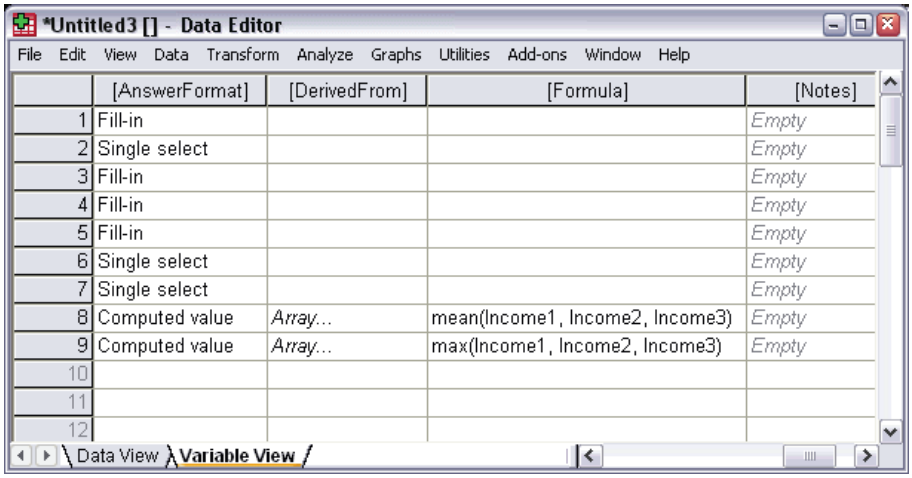
Visualizza attributo nell'Editor dei dati. Visualizza l'attributo in Visualizzazione variabili di Editor dei dati. Per informazioni sul controllo della visualizzazione degli attributi personalizzati, vedere Visualizzazione e modifica di attributi di variabile personalizzati di seguito.

Visualizza Elenco definito di attributi. Visualizza un elenco di attributi personalizzati già definiti per l'insieme dei dati. I nomi di attributo che iniziano con il simbolo dollaro ($) appartengono ad attributi riservati e non possono essere modificati.

Visualizzazione e modifica di attributi di variabile personalizzati

Gli attributi di variabile personalizzati possono essere visualizzati e modificati in Visualizzazione variabili di Editor dei dati.

Figura 5-7
Attributi di variabile personalizzati visualizzati in Visualizzazione variabili


*Untitled3 [] - Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
[AnswerFormat] [DerivedFrom] [Formula] [Notes]
1 Fill-in Empty
2 Single select Empty
3 Fill-in Empty
4 Fill-in Empty
5 Fill-in Empty
6 Single select Empty
7 Single select Empty
8 Computed value Array... mean(Income1, Income2, Income3) Empty
9 Computed value Array... max(Income1, Income2, Income3) Empty
10
11
12
Data View Variable View

- I nomi di attributi di variabile personalizzati si presentano racchiusi tra parentesi.
- I nomi di attributo che iniziano con il simbolo dollaro ($) appartengono ad attributi riservati e non possono essere modificati.
- Una cella vuota indica che l'attributo non esiste per quella variabile; il testo *Vuoto* che appare in una cella indica che l'attributo esiste per quella variabile ma che nessun valore è stato assegnato all'attributo per quella variabile. Digitare un testo nella cella per valorizzare l'attributo per quella variabile.
- Il testo *Matrice...* visualizzato in una cella indica che si tratta di una matrice di attributi, ossia un attributo contenente più valori. Fare clic sul pulsante della cella per visualizzare l'elenco dei valori.

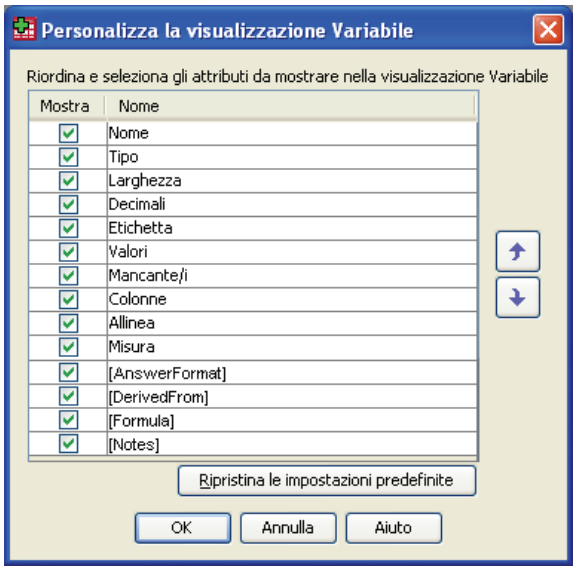
Per visualizzare e modificare attributi di variabile personalizzati

▶ Dai menu della Visualizzazione variabili scegliere:

Visualizza
  Personalizza visualizzazione Variabile...

▶ Selezionare gli attributi di variabile personalizzati da visualizzare. Gli attributi di variabile personalizzati si presentano racchiusi tra parentesi.

Figura 5-8
Personalizza la visualizzazione Variabile

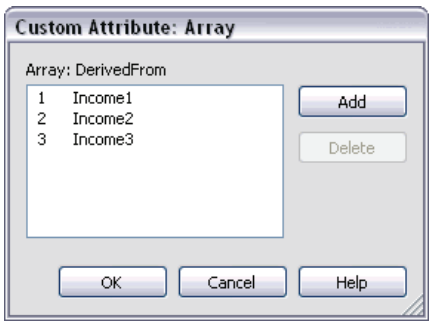


Una volta che gli attributi vengono mostrati nella Visualizzazione variabili, è possibile modificarli direttamente nell'Editor dei dati.

Matrici di attributi di variabile

Il testo *Matrice...* visualizzato in una cella per l'attributo di variabile personalizzato nella Visualizzazione variabili o nella finestra di dialogo Proprietà in Definisci proprietà variabili indica che si tratta di una **matrice di attributi**, un attributo contenente più valori. Ad esempio, è possibile disporre di una matrice di attributi che identifica tutte le variabili di origine usate per il calcolo di una variabile derivata. Fare clic sul pulsante della cella per visualizzare e modificare l'elenco dei valori.

Figura 5-9
Finestra di dialogo Attributo personalizzato: matrice



Personalizzazione della Visualizzazione variabili

È possibile utilizzare Personalizza la visualizzazione Variabile per controllare quali attributi vengono mostrati nella Visualizzazione variabili (ad esempio nome, tipo, etichetta) e l'ordine in cui vengono mostrati.

- Gli attributi di variabile personalizzati associati all'insieme di dati si presentano racchiusi tra parentesi. Per ulteriori informazioni, vedere Creazione di attributi di variabile personalizzati a pag. 88.
- Le impostazioni di visualizzazione personalizzate vengono salvate con file di dati in formato SPSS Statistics.
- È inoltre possibile controllare la visualizzazione e l'ordine predefiniti degli attributi nella Visualizzazione variabili. Per ulteriori informazioni, vedere Modifica della Visualizzazione variabili predefinita in Capitolo 48 a pag. 527.

Per personalizzare la Visualizzazione variabili

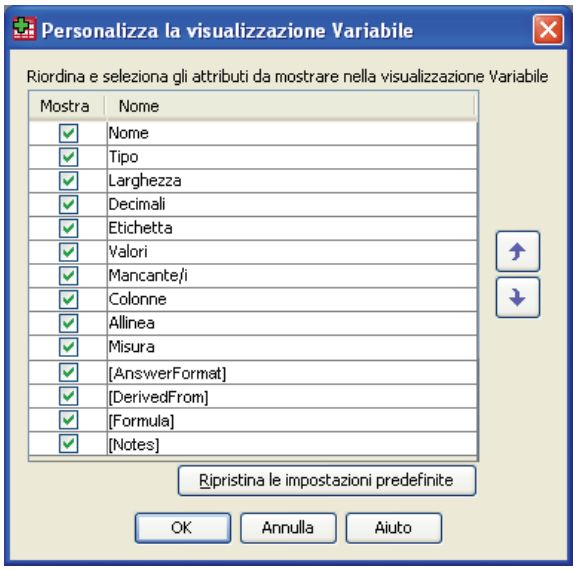
▶ Dai menu della Visualizzazione variabili scegliere:

Visualizza
Personalizza visualizzazione Variabile...

▶ Selezionare gli attributi di variabile da visualizzare.

▶ Utilizzare i pulsanti di spostamento verso l'alto e verso il basso per cambiare l'ordine di visualizzazione degli attributi.

Figura 5-10
Finestra di dialogo Personalizza la visualizzazione Variabile



Ripristina valori predefiniti. Consente di applicare le impostazioni predefinite per la visualizzazione e l'ordinamento.

Controllo ortografico

Etichette delle variabili e dei valori

Per controllare l'ortografia delle etichette delle variabili e delle etichette dei valori:

- Selezionare la scheda Visualizzazione variabili nella finestra dell'Editor dei dati.
- Fare clic con il pulsante destro del mouse sulla colonna *Etichette* o *Valori* e dal menu di scelta rapida scegliere:
  Ortografia

  o
- Dai menu della Visualizzazione variabili scegliere:
  Strumenti
  Ortografia

  o
- Nella finestra di dialogo Etichette dei valori, fare clic su Ortografia. Questa operazione limita il controllo ortografico alle sole etichette dei valori di una particolare variabile.

Il controllo ortografico è limitato alle sole etichette delle variabili e dei valori nella Visualizzazione variabili dell'Editor dei dati.

Valori di dati stringa

Per controllare l'ortografia dei valori di dati stringa:

- Selezionare la scheda Visualizzazione dati dell'Editor dei dati.
- Se si desidera, è possibile selezionare una o più variabili (colonne) da controllare. Per selezionare una variabile, fare clic sul nome corrispondente alla base della colonna.
- Dai menu, scegliere:
  Strumenti
  Ortografia
  - Se non vi sono variabili selezionate nella Visualizzazione dati, verranno controllate tutte le variabili stringa.
  - Se l'insieme di dati non contiene variabili stringa o se nessuna delle variabili selezionate è una variabile stringa, l'opzione Controllo ortografico nel menu Strumenti è disabilitato.

Inserimento di dati

In Visualizzazione dati è possibile immettere i dati direttamente nell'Editor dei dati. I dati possono essere inseriti in qualsiasi ordine. I dati possono essere inseriti per caso o per variabile nelle aree selezionate o nelle singole celle.

- La cella attiva viene evidenziata.

- Il nome della variabile e il numero di riga della cella attiva verranno visualizzati nell'angolo in alto a sinistra dell'Editor dei dati.
- Quando si seleziona una cella e si inserisce un valore, questo verrà visualizzato nell'editor delle celle nella parte superiore dell'Editor dei dati.
- I valori dei dati verranno registrati solo quando si preme Invio o si seleziona un'altra cella.
- Per inserire dati diversi da semplici dati numerici, è innanzitutto necessario definire il tipo di variabile.

Se si inserisce un valore in una colonna vuota, verrà creata automaticamente una nuova variabile a cui verrà assegnato un nome.

Per inserire dati numerici

▶ Selezionare una cella in Visualizzazione dati.

▶ Inserire il valore. Il valore viene visualizzato nell'editor celle nella parte superiore dell'Editor dei dati.

▶ Per registrare il valore, premere Invio oppure selezionare un'altra cella.

Per inserire dati non numerici

▶ Fare doppio clic su un nome di variabile nella parte superiore della colonna in Visualizzazione dati oppure fare clic sulla scheda Visualizzazione variabili.

▶ Fare clic sul pulsante nella cella *Tipo* per la variabile.

▶ Selezionare il tipo di dati desiderato nella finestra di dialogo Tipo di variabile.

▶ Fare clic su OK.

▶ Fare doppio clic sul numero di riga oppure sulla scheda Visualizzazione dati.

▶ Inserire i dati nella colonna della nuova variabile definita.

Per utilizzare etichette dei valori per l'inserimento dei dati

▶ Se in Visualizzazione dati non sono visualizzate le etichette dei valori, scegliere:
Visualizza
Etichette dei valori

▶ Fare clic sulla cella in cui immettere il valore.

▶ Selezionare un'etichetta dall'elenco a discesa.

Il valore verrà inserito e l'etichetta verrà visualizzata nella cella.

Nota: Questo processo funziona solo se per la variabile sono state definite etichette di valori.

Limitazioni relative ai valori

Il tipo di valore che può essere inserito nella cella in Visualizzazione dati viene determinato in base al tipo di variabile e alla larghezza definiti.

- Se si digita un carattere non consentito dal tipo di variabile definito, il carattere non verrà inserito.
- Nelle variabili stringa non è consentito inserire un numero di caratteri superiore a quello definito.
- Nelle variabili numeriche è possibile inserire valori interi maggiori della larghezza definita, ma nell'Editor dei dati verranno visualizzati una notazione scientifica oppure una parte del valore seguita dai puntini (...) ad indicare che il valore supera la larghezza definita. Per visualizzare il valore nella cella, è necessario modificare la larghezza definita per la variabile.

  Nota: la modifica della larghezza della colonna non comporta la modifica della lunghezza della variabile.

Modifica dei dati nella Visualizzazione dati

Con l'Editor dei dati è possibile modificare i valori dei dati in Visualizzazione dati in diversi modi. È possibile:

- Modificare i valori dei dati
- Tagliare, copiare e incollare i valori dei dati
- Aggiungere ed eliminare casi
- Aggiungere ed eliminare variabili
- Modificare l'ordine delle variabili

Sostituzione o modifica di valori

Per eliminare il vecchio valore e inserire un nuovo valore

▶ In Visualizzazione dati, fare doppio clic sulla cella. Il valore della cella verrà visualizzato nell'editor delle celle.

▶ Modificare il valore direttamente nella cella o nell'editor delle celle.

▶ Premere Invio oppure selezionare un'altra cella per registrare il nuovo valore.

Tagliare, copiare e incollare valori

Nell'Editor dei dati è possibile tagliare, copiare e incollare i valori delle singole celle oppure gruppi di valori. È possibile:

- Spostare o copiare il valore di una cella in un'altra cella
- Spostare o copiare il valore di una cella in un gruppo di celle
- Spostare o copiare i valori di un singolo caso (riga) in più casi

- Spostare o copiare i valori di una singola variabile (colonna) in più variabili
- Spostare o copiare i valori di un gruppo di celle in un altro gruppo di celle

Conversione di dati per i valori incollati nella Visualizzazione dati

Se i tipi di variabile definiti delle celle di origine non coincidono con quelli delle celle di destinazione, verrà eseguito un tentativo di conversione del valore. Se la conversione non è possibile, nella cella di destinazione verrà inserito il valore mancante di sistema.

Conversione di valori numerici o di date in stringhe. I formati numerici (ad esempio numerico, dollaro, punto o virgola) e della data verranno convertiti in stringhe se incollati nella cella di una variabile stringa. Il valore stringa corrisponde al valore numerico visualizzato nella cella. Per una variabile in formato dollaro, il segno di dollaro visualizzato diventa parte del valore stringa. I valori di larghezza superiore a quella definita per la variabile stringa verranno troncati.

Conversione di stringhe in valori numerici o date. Le stringhe contenenti caratteri consentiti per il formato numerico o della data della cella di destinazione verranno convertiti nel corrispondente valore numerico o data. Il valore stringa 25/12/91, ad esempio, viene convertito in una data valida se il tipo di formato della cella di destinazione è uno dei formati giorno-mese-anno; se invece quest'ultimo è uno dei formati mese-giorno-anno, il valore verrà convertito in valore mancante di sistema.

Conversione di date in valori numerici. I valori relativi alla data e all'ora vengono convertiti in un numero di secondi se la cella di destinazione corrisponde a un formato numerico (ad esempio, numerico, dollaro, punto o virgola). Poiché le date vengono memorizzate internamente come il numero di secondi trascorsi dal 14 ottobre 1582, la conversione delle date in valori numerici può produrre risultati composti da un numero considerevole di cifre. La data 29/10/91, ad esempio, viene convertita nel valore numerico 12.908.073.600.

Conversione di valori numerici in date od ore. I valori numerici vengono convertiti in date o orari se il valore rappresenta un numero di secondi che può produrre una data o un orario valido. Per quanto riguarda le date, i valori numerici inferiori a 86.400 verranno convertiti nel valore mancante di sistema.

Inserimento di nuovi casi

Se si immettono dati in una cella o in una riga vuota verrà creato automaticamente un nuovo caso. Verrà inserito automaticamente il valore mancante di sistema per tutte le altre variabili del caso specifico. Se tra il nuovo caso e i casi esistenti sono interposti spazi vuoti, anche le righe vuote diventeranno nuovi casi con il valore mancante di sistema specificato per tutte le variabili. È inoltre possibile inserire nuovi casi tra casi esistenti.

Per inserire un nuovo caso tra casi esistenti

▶ In Visualizzazione dati, selezionare una cella qualsiasi nel caso (riga) sottostante la posizione in cui si desidera inserire un nuovo caso.

- Dai menu, scegliere:
  Modifica
  Inserisci casi

  Verrà inserita una nuova riga per il caso e in tutte le variabili verrà immesso il valore mancante di sistema.

Inserimento di nuove variabili

L'immissione di dati in una colonna vuota di Visualizzazione dati o in una riga vuota di Visualizzazione variabili comporta la creazione automatica di una nuova variabile con un nome predefinito (il prefisso *var* e un numero progressivo) e di un tipo di formato dei dati predefinito (numerico). Per tutti i casi della nuova variabile verrà inserito automaticamente il valore mancante di sistema. Se in Visualizzazione dati o in Visualizzazione variabili vi sono rispettivamente delle colonne o delle righe vuote comprese tra la nuova variabile e quelle esistenti, anche queste divengono nuove variabili con il valore mancante di sistema per tutti i casi. È inoltre possibile inserire una nuova variabile tra le variabili esistenti.

Per inserire una nuova variabile tra le variabili esistenti

- Selezionare qualsiasi cella della variabile a destra (Visualizzazione dati) o sotto (Visualizzazione variabili) la posizione in cui inserire la nuova variabile.
- Dai menu, scegliere:
  Modifica
  Inserisci variabile

  Verrà inserita una nuova variabile con il valore mancante di sistema per tutti i casi.

Per spostare le variabili

- Per selezionare la variabile, fare clic sul nome della variabile in Visualizzazione dati o sul numero di riga della variabile in Visualizzazione variabili.
- Trascinare la variabile nella nuova posizione.
- Per inserire la variabile tra due variabili esistenti: in Visualizzazione dati, inserire la variabile nella colonna variabili a destra del punto in cui si desidera inserirla oppure in Visualizzazione variabili inserirla nella riga variabili sotto al punto in cui si desidera inserirla.

Per modificare il tipo di dati

È possibile modificare il tipo di dati di una variabile in qualunque momento tramite la finestra di dialogo Tipo di variabile in Visualizzazione variabili. L'Editor dei dati tenta di convertire i valori esistenti nel nuovo tipo. Se la conversione non è possibile, verrà assegnato il valore mancante di sistema. Le regole di conversione sono analoghe a quelle utilizzate per incollare valori in una variabile con un formato diverso. Se la modifica del formato dei dati comporta la perdita delle specifiche dei valori mancanti o delle etichette dei valori, verrà visualizzata una finestra di messaggio in cui viene richiesto se procedere alla modifica o annullarla.

Ricerca di casi, variabili o assegnazioni

La finestra di dialogo Vai a trova il numero di caso specificato (riga) o nome di variabile nell'Editor dei dati.

Casi

- Per i casi, dai menu scegliere:

  Modifica
  Vai al caso...

- Immettere un valore intero che rappresenta il numero di riga corrente nella Visualizzazione dati.

  Nota: il numero di riga corrente per un determinato caso può cambiare in seguito all'ordinamento e ad altre azioni.

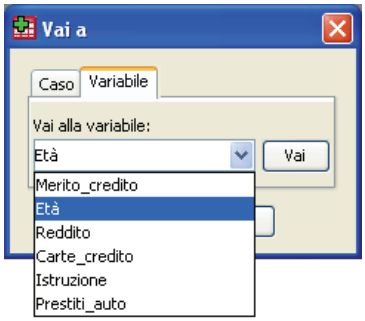
Variabili

- Per le variabili, dai menu scegliere:

  Modifica
  Vai alla variabile...

- Immettere il nome della variabile o selezionarla dall'elenco a discesa.

Figura 5-11
Finestra di dialogo Vai a



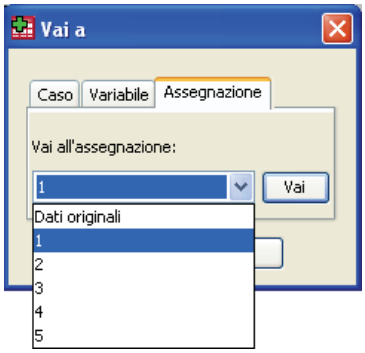
Assegnazioni

- Dai menu, scegliere:

  Modifica
  Vai all'assegnazione...

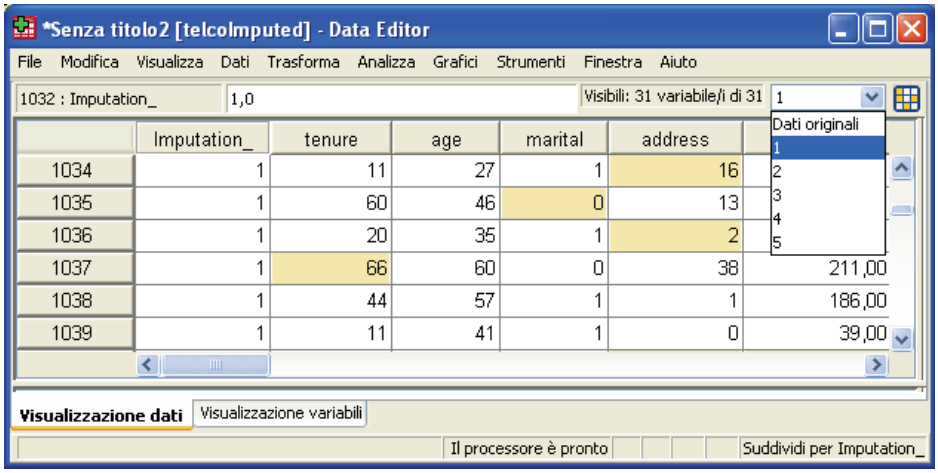
- Selezionare l'assegnazione (o Dati originali) dall'elenco a discesa.

Figura 5-12
Finestra di dialogo Vai a



In alternativa, è possibile selezionare l'assegnazione dall'elenco a discesa nella barra di modifica.

Figura 5-13
Editor dei dati con contrassegni di assegnazione attivati



La posizione relativa dei casi viene mantenuta quando si selezionano le assegnazioni. Ad esempio, se ci sono 1000 casi nell'insieme di dati originale, il caso 1034, il trentaquattresimo nella prima assegnazione, viene visualizzato nella parte superiore della griglia. Se si seleziona l'assegnazione 2 nell'elenco a discesa, il caso 2034, il trentaquattresimo dell'assegnazione 2, viene visualizzato nella parte superiore della griglia. Se si seleziona Dati originali nell'elenco a discesa, il caso 34 viene visualizzato nella parte superiore della griglia. Quando si passa da un'assegnazione all'altra, viene mantenuta anche la posizione delle colonne, in modo da agevolare il confronto dei valori tra le assegnazioni.

Ricerca e sostituzione dei valori dei dati e degli attributi

Per trovare e/o sostituire i valori dei dati nella Visualizzazione dati o i valori degli attributi nella Visualizzazione variabili:

- Fare clic su una cella nella colonna che si desidera ricercare. Le operazioni di ricerca e sostituzione dei valori sono limitate a una sola colonna.

▶ Dai menu, scegliere:

Modifica
  Trova

o

Modifica
  Sostituisci

Visualizzazione dati

- Nella Visualizzazione dati non è possibile ricercare i dati verso l'alto. La direzione di ricerca è sempre verso il basso.
- Per i valori di data e ora, i valori formattati vengono ricercati così come vengono mostrati nella Visualizzazione dati. Ad esempio, una data visualizzata come 10/28/2007 non verrà trovata da un'operazione di ricerca in cui la data sia specificata in formato 10-28-2007.
- Per le variabili numeriche, Contiene, Inizia con e Termina con ricercano i valori formattati. Ad esempio, con l'opzione Inizia con, un valore di ricerca $123 per una variabile in formato dollaro troverà sia $123.00 che $123.40 ma non $1,234. Con l'opzione Cella intera, il valore di ricerca può essere formattato o non formattato (formato numerico semplice F), ma vengono confrontati solo i valori numerici esatti (in base alla precisione mostrata nell'Editor dei dati).
- Se le etichette dei valori vengono visualizzate per la colonna della variabile selezionata, viene ricercato il testo dell'etichetta e non il valore dei dati sottostanti e non è possibile sostituire il testo dell'etichetta.

Visualizzazione variabili

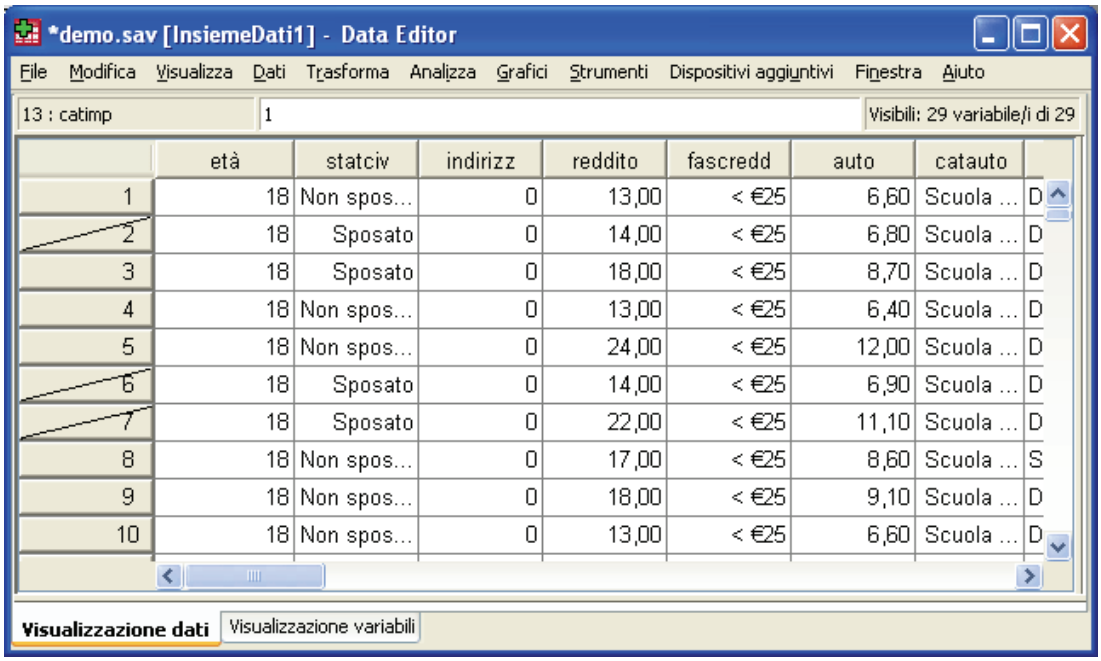
- Trova è disponibile soltanto per le colonne *Nome*, *Etichetta*, *Valori*, *Mancante* e per le colonne degli attributi delle variabili personalizzati.
- Sostituisci è disponibile solo per le colonne *Etichetta*, *Valori* e per le colonne degli attributi personalizzati.
- Nella colonna *Valori* (etichette dei valori), la stringa di ricerca può corrispondere al valore dei dati o a un'etichetta del valore.

  Nota: la sostituzione del valore dei dati eliminerà qualsiasi etichetta di valore precedente associata a tale valore.

Criteri di selezione dei casi nell'Editor dei dati

Se si è selezionato un sottoinsieme di casi senza eliminare i casi non selezionati, nell'Editor dei dati questi ultimi verranno contrassegnati da una linea diagonale (barra) sopra il numero di riga.

Figura 5-14
Casi filtrati nell'Editor dei dati


*demo.sav [InsiemeDati1] - Data Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntivi Finestra Aiuto
13 : catimp
1
Visibili: 29 variabile/i di 29
età statciv indirizz reddito fascredd auto catauto
1 18 Non spos... 0 13,00 < €25 6,60 Scuola ... D
2 18 Sposato 0 14,00 < €25 6,80 Scuola ... D
3 18 Sposato 0 18,00 < €25 8,70 Scuola ... D
4 18 Non spos... 0 13,00 < €25 6,40 Scuola ... D
5 18 Non spos... 0 24,00 < €25 12,00 Scuola ... D
6 18 Sposato 0 14,00 < €25 6,90 Scuola ... D
7 18 Sposato 0 22,00 < €25 11,10 Scuola ... D
8 18 Non spos... 0 17,00 < €25 8,60 Scuola ... S
9 18 Non spos... 0 18,00 < €25 9,10 Scuola ... D
10 18 Non spos... 0 13,00 < €25 6,60 Scuola ... D
Visualizzazione dati
Visualizzazione variabili

Editor dei dati: Opzioni di visualizzazione

Nel menu Visualizza sono disponibili diverse opzioni di visualizzazione per l'Editor dei dati:

Caratteri. Questa opzione consente di controllare le caratteristiche dei caratteri per la visualizzazione dei dati.

Griglia. Questa opzione attiva o disattiva la visualizzazione delle linee della griglia.

Etichette dei valori. Questa opzione consente di visualizzare alternativamente i dati effettivi e le etichette descrittive dei valori definite dall'utente. Ciò è possibile solo in Visualizzazione dati.

Uso delle viste multiple

Nella Visualizzazione dati è possibile creare viste multiple (riquadri) utilizzando i divisori situati sotto la barra di scorrimento orizzontale e a destra della barra di scorrimento verticale.

È possibile anche utilizzare il menu Finestra per inserire e rimuovere i divisori dei riquadri. Per inserire i divisori:

▶ Nei menu della Visualizzazione dati scegliere:

Finestra
  Dividi

I divisori vengono inseriti sopra e a sinistra della cella selezionata.

- Se si seleziona la cella in alto a sinistra, i divisori inseriti dividono la vista corrente a circa metà, sia orizzontalmente che verticalmente.

- Se si seleziona una cella diversa dalla prima cella della prima colonna, viene inserito un divisore di riquadri orizzontale sopra la cella selezionata.
- Se si seleziona una cella diversa dalla prima cella della prima riga, viene inserito un divisore di riquadri verticale a sinistra della cella selezionata.

Stampa dall'Editor dei dati

I file di dati vengono stampati esattamente come vengono visualizzati.

- Vengono stampati i dati presenti nella vista correntemente visualizzata. In Visualizzazione dati, vengono stampati i dati. In Visualizzazione variabili, vengono stampate le informazioni di definizione dei dati.
- Le griglie vengono stampate se visualizzate nella vista selezionata.
- Le etichette dei valori vengono stampate in Visualizzazione dati se sono visualizzate. In caso contrario, vengono stampati i valori dei dati effettivi.

Utilizzare il menu Visualizza nell'Editor dei dati per visualizzare o nascondere la griglia e impostare la visualizzazione dei valori o delle etichette dei valori.

Per stampare il contenuto dell'Editor dei dati

▶ Attivare la finestra dell'Editor dei dati.

▶ Selezionare la scheda relativa alla visualizzazione da stampare.

▶ Dai menu, scegliere:

File
Stampa...

Capitolo

6

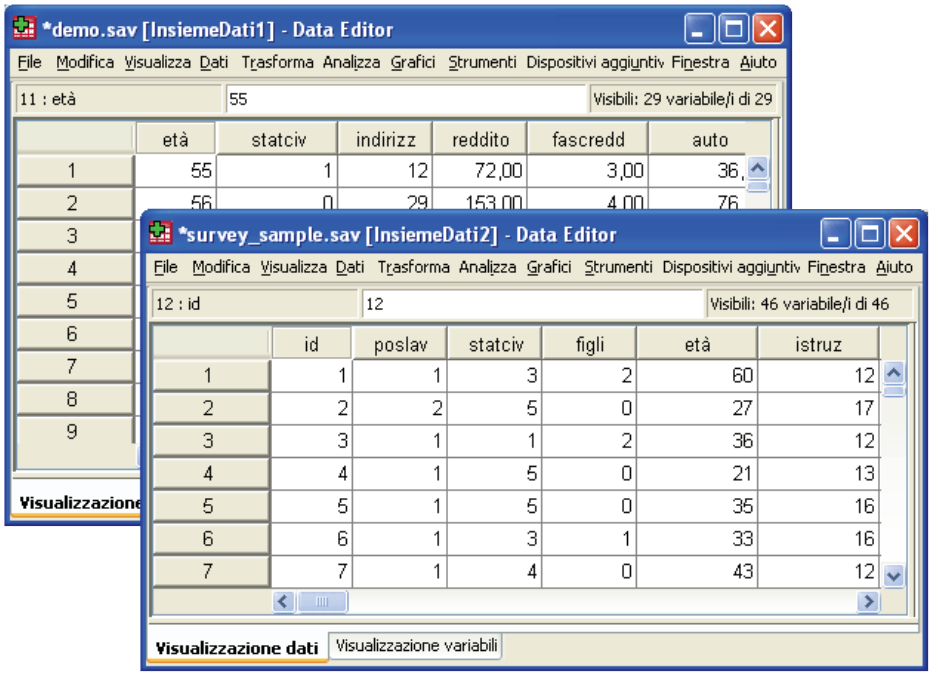
Utilizzo di più sorgenti dati

A partire dalla versione 14.0, è consentito tenere aperte più sorgenti dati contemporaneamente. Ciò permette agevolmente di:

- Passare da una sorgente dati a un'altra.
- Confrontare i contenuti di più sorgenti dati.
- Copiare e incollare i dati tra le sorgenti dati.
- Creare più sottoinsiemi di casi e/o di variabili per l'analisi.
- Unire più sorgenti dati di formati dati diversi (ad esempio fogli di calcolo, database e dati di testo) senza dover salvare prima ciascuna sorgente dati.

Gestione di più sorgenti di dati

Figura 6-1
Due sorgenti di dati aperte contemporaneamente

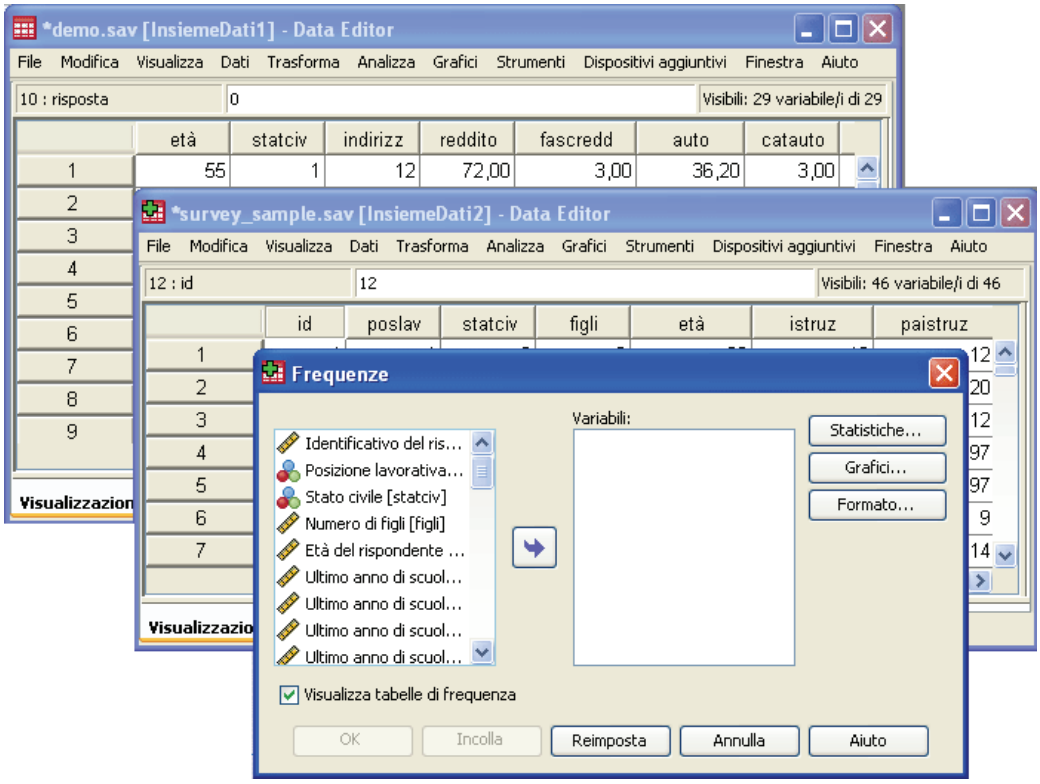

*demo.sav [InsiemeDati1] - Data Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntiv Finestra Aiuto
11 : età
55
Visibili: 29 variabile/i di 29
età statciv indirizz reddito fascredd auto
1 55 1 12 72,00 3,00 36,
2 56 0 29 153,00 4,00 76
3
4
5
6
7
8
9
Visualizzazione
*survey_sample.sav [InsiemeDati2] - Data Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntiv Finestra Aiuto
12 : id
12
Visibili: 46 variabile/i di 46
id poslav statciv figli età istruz
1 1 1 3 2 60 12
2 2 2 5 0 27 17
3 3 1 1 2 36 12
4 4 1 5 0 21 13
5 5 1 5 0 35 16
6 6 1 3 1 33 16
7 7 1 4 0 43 12
Visualizzazione dati
Visualizzazione variabili

Per impostazione predefinita, le sorgenti dati aperte vengono visualizzate in una nuova finestra dell'Editor dei dati. Vedere Opzioni generali per informazioni sulla modifica del comportamento predefinito che consente di visualizzare un solo insieme di dati alla volta in un'unica finestra dell'Editor dei dati.

- Tutte le sorgenti di dati già aperte rimangono tali e possono essere usate.
- La sorgente di dati aperte diventa automaticamente il **insieme di dati attivo**.
- Per modificare l'insieme di dati attivo, è sufficiente fare clic in qualsiasi punto della finestra dell'Editor dei dati della sorgente di dati da usare oppure selezionare la finestra Editor dei dati per la sorgente di dati dal menu Finestra.

- Solo le variabili dell'insieme di dati attivo possono essere usate per l'analisi.

Figura 6-2
Elenco di variabili contenente le variabili dell'insieme di dati attivo



- L'insieme di dati attivo non può essere modificato se è aperta una finestra di dialogo che ha accesso ai dati (ciò include anche le finestre di dialogo che visualizzano gli elenchi di variabili).
- È sempre necessario aprire almeno una finestra dell'Editor dei dati per sessione. Quando si chiude l'ultima finestra aperta dell'Editor dei dati, SPSS Statistics la chiude automaticamente chiedendo all'utente se desidera salvare le modifiche.

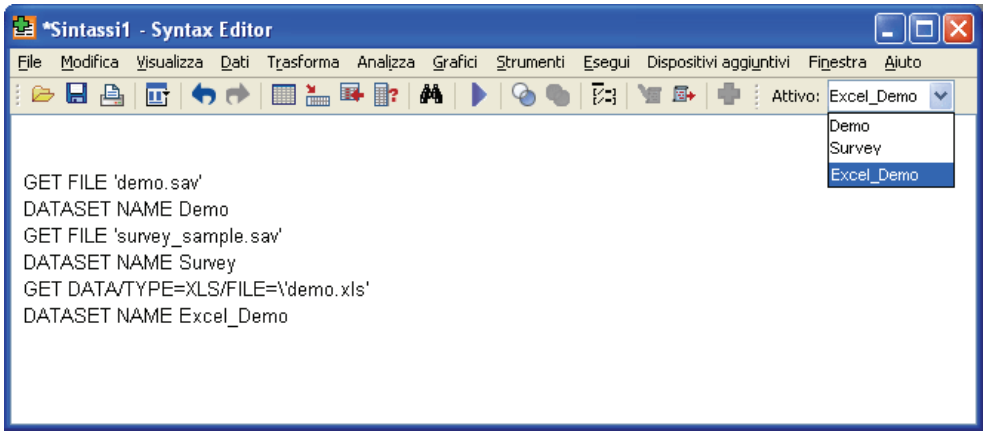
Utilizzo di più insiemi di dati nella sintassi dei comandi

Se si utilizza la sintassi dei comandi per aprire le sorgenti dati (ad esempio, `GET FILE`, `GET DATA`), è necessario utilizzare il comando `DATASET NAME` per denominare esplicitamente ciascun insieme di dati per poter tenere contemporaneamente aperte più sorgenti dati.

Quando si utilizza la sintassi dei comandi, il nome dell'insieme di dati attivo viene visualizzato sulla barra degli strumenti della finestra della sintassi. Tutte le azioni riportate di seguito consentono di modificare l'insieme di dati attivo:

- Utilizzare il comando `DATASET ACTIVATE`.
- Fare clic in un punto qualsiasi della finestra dell'Editor dei dati di un insieme di dati.
- Selezionare un nome di insieme di dati dalla barra degli strumenti nella finestra della sintassi.

Figura 6-3
Insiemi di dati aperti visualizzati sulla barra degli strumenti della finestra della sintassi



Copia e inserimento di informazioni in più insiemi di dati

È possibile copiare gli attributi delle definizioni dei dati e delle variabili da un insieme di dati a un altro eseguendo le stesse operazioni utilizzabili all'interno dei singoli file di dati.

- Se si copiano e si incollano le celle dati selezionati in Visualizzazione dei dati, vengono incollati solo i valori dei dati senza gli attributi delle definizioni della variabili.
- Se si copia e si incolla una variabile intera nella Visualizzazione dei dati selezionando il nome della variabile nella parte superiore della colonna, vengono incollati tutti i dati e tutti gli attributi delle definizioni delle variabili per la variabile in questione.
- Se si copiano e si incollano gli attributi delle definizioni delle variabili o le variabili intere nella Visualizzazione variabili, vengono incollati gli attributi selezionati (o la definizione completa della variabile) ma non i valori dati.

Ridenominazione degli insiemi di dati

Quando si apre una sorgente dati tramite i menu e le finestre di dialogo, a ciascuna sorgente viene automaticamente assegnato il nome di file dati *DataSetn*, dove *n* è il valore intero sequenziale. Quando si apre una sorgente dati tramite la sintassi del comando, non viene assegnato alcun nome di insieme di dati a meno che non se ne specifici uno con il comando `DATASET NAME`. Per assegnare un nome di insieme di dati più descrittivo:

▶ Dai menu della finestra Editor dei dati relativo all'insieme di dati di cui si vuole modificare il nome selezionare:

File
  Rinomina insieme dati...

▶ Immettere il nome di un nuovo insieme di dati conforme alle regole di denominazione delle variabili. Per ulteriori informazioni, vedere Nomi delle variabili in Capitolo 5 a pag. 79.

Soppressione di più insiemi di dati

Se si preferisce avere a disposizione un solo insieme di dati alla volta e si desidera sopprimere la funzione relativa a più insiemi di dati:

- Dai menu, scegliere:

  Modifica
  Opzioni...

- Fare clic sulla scheda Generale.

Selezionare Apri un solo insieme di dati alla volta.

Per ulteriori informazioni, vedere Opzioni generali in Capitolo 48 a pag. 522.

Capitolo
7

Preparazione dei dati

Dopo avere aperto un file di dati o immesso dei dati nell'Editor, è possibile creare report, grafici e analisi senza ulteriori operazioni preliminari. Sono tuttavia disponibili alcune opzioni aggiuntive per la preparazione dei dati che potrebbero rivelarsi utili, ad esempio:

- assegnazione delle proprietà delle variabili che descrivono i dati e determinazione delle modalità di utilizzo di determinati valori.
- Identificazione dei casi che potrebbero contenere informazioni duplicate ed esclusione di tali casi dalle analisi o eliminazione degli stessi dal file di dati.
- Creazione di nuove variabili con categorie distinte che rappresentano gli intervalli di valori delle variabili con un numero elevato di possibili valori.

Proprietà variabili

I dati immessi nella Visualizzazione dati dell'Editor dei dati oppure letti da un formato di file esterno, ad esempio un foglio elettronico di Excel o un file di dati di testo, non dispongono di alcune proprietà di variabili che potrebbero essere molto utili, quali:

- Definizione delle etichette di valore descrittive per codici numerici (per esempio, 0 = *Maschio* e 1 = *Femmina*).
- Identificazione dei codici di valori mancanti (per esempio, 99 = *Non applicabile*).
- Assegnazione del livello di misurazione (nominale, ordinale o di scala).

È possibile assegnare tutte queste ed altre proprietà di variabili nella Visualizzazione variabili nell'Editor dei dati. Per agevolare l'esecuzione di tale procedura è possibile ricorrere anche ai seguenti strumenti:

- **Definisci proprietà variabili** consente di definire le etichette dei valori descrittivi e i valori mancanti. Si tratta di uno strumento particolarmente utile per i dati categoriali con codici numerici utilizzati per i valori delle categorie.
- **Copia proprietà dei dati** consente di utilizzare un file di dati di SPSS Statistics esistente come modello per le proprietà di file e variabili nel file di dati corrente. Si tratta di uno strumento particolarmente utile se si utilizzano spesso file di dati esterni con contenuto simile, come i report mensili in formato Excel.

Definizione di proprietà delle variabili

Definisci proprietà variabili semplifica il processo di creazione di etichette di valori descrittive per variabili categoriali (nominali, ordinali). Definisci proprietà variabili:

- Esamina i valori effettivi dei dati ed elenca tutti i valori univoci dei dati per ogni variabile selezionata.
- Identifica i valori senza etichetta e fornisce una funzione per l'aggiunta automatica dell'etichetta.
- Consente di copiare etichette di valori definite da un'altra variabile nella variabile selezionata o dalla variabile selezionata in più variabili aggiuntive.

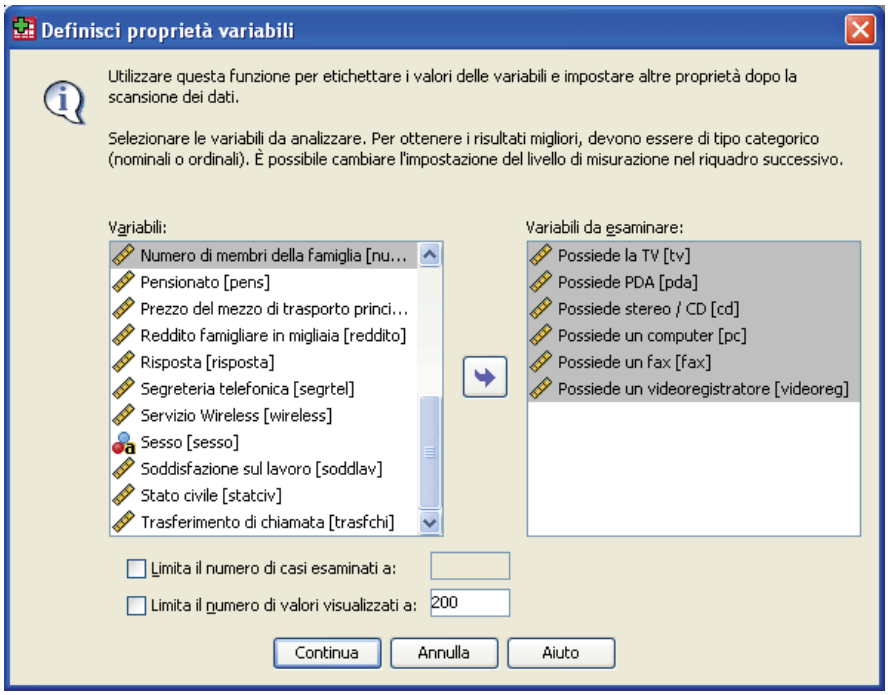
Nota: per utilizzare Definisci proprietà variabili senza eseguire la scansione dei casi, specificare 0 per il numero di casi da esaminare.

Per definire le proprietà delle variabili

▶ Dai menu, scegliere:

Dati
Definisci proprietà variabili...

Figura 7-1
Finestra di dialogo iniziale per la selezione delle variabili da definire

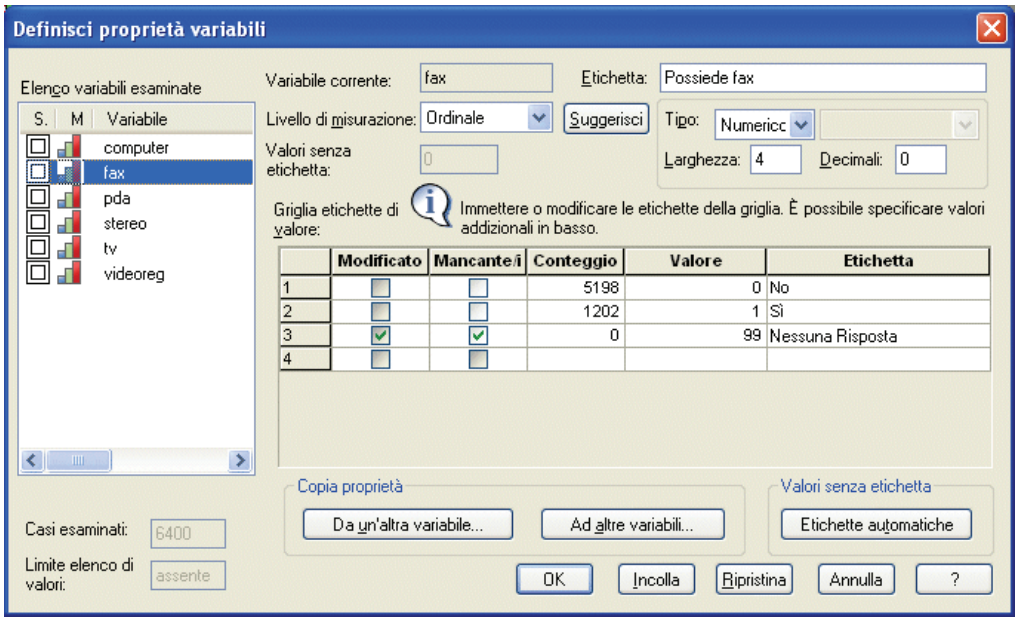

Definisci proprietà variabili
Utilizzare questa funzione per etichettare i valori delle variabili e impostare altre proprietà dopo la scansione dei dati.
Selezionare le variabili da analizzare. Per ottenere i risultati migliori, devono essere di tipo categorico (nominali o ordinali). È possibile cambiare l'impostazione del livello di misurazione nel riquadro successivo.
Variabili:
Numero di membri della famiglia [nu...
Pensionato [pens]
Prezzo del mezzo di trasporto princi...
Reddito famigliare in migliaia [reddito]
Risposta [risposta]
Segreteria telefonica [segrtel]
Servizio Wireless [wireless]
Sesso [sesso]
Soddisfazione sul lavoro [soddlav]
Stato civile [statciv]
Trasferimento di chiamata [trasfchi]
Variabili da esaminare:
Possiede la TV [tv]
Possiede PDA [pda]
Possiede stereo / CD [cd]
Possiede un computer [pc]
Possiede un fax [fax]
Possiede un videoregistratore [videoreg]
Limita il numero di casi esaminati a:
Limita il numero di valori visualizzati a:
200
Continua
Annulla
Aiuto

▶ Selezionare le variabili stringa o numeriche per le quali si desidera creare etichette di valori oppure definire o modificare altre proprietà delle variabili, ad esempio valori mancanti o etichette di variabile descrittive.

- Specificare il numero di casi da esaminare per generare l'elenco di valori univoci. Questo risulta particolarmente utile per i file di dati con un numero elevato di casi per cui una scansione del file di dati completo potrebbe richiedere una notevole quantità di tempo.
- Specificare un limite superiore per il numero di valori univoci da visualizzare. Ciò è utile soprattutto per impedire che vengano elencati centinaia, migliaia o anche milioni di valori per variabili di scala (rapporto, intervallo continuo).
- Fare clic su Continua per aprire la finestra di dialogo Definisci proprietà variabili.
- Selezionare una variabile per cui si desidera creare etichette di valori oppure definire o modificare altre proprietà delle variabili.
- Immettere il testo dell'etichetta per tutti i valori senza etichetta visualizzati nella Griglia etichette di valore.
- Se si desidera creare etichette di valore per valori non visualizzati, è possibile immettere tali valori nella colonna *Valore* che si trova al di sotto dell'ultimo valore esaminato.
- Ripetere l'operazione per ogni variabile elencata per cui si desidera creare etichette di valori.
- Fare clic su OK per applicare le etichette di valore e le altre proprietà delle variabili definite.

Definizione di etichette di valori e di altre proprietà delle variabili

Figura 7-2
Definisci proprietà variabili, finestra di dialogo principale



La finestra di dialogo principale Definisci proprietà variabili include le seguenti informazioni sulle variabili esaminate:

Elenco variabili esaminate. Un segno di spunta nella colonna *Senza etichetta* (*U.*) indica che la variabile contrassegnata contiene valori senza etichetta assegnata.

Per ordinare l'elenco delle variabili in modo che le variabili con valori senza etichetta vengano visualizzate all'inizio dell'elenco:

▶ Fare clic sull'intestazione della colonna *Senza etichetta* nell'Elenco variabili esaminate.

È inoltre possibile ordinare l'elenco per nome di variabile o livello di misurazione facendo clic sull'intestazione di colonna corrispondente nell'Elenco variabili esaminate.

Griglia etichette di valore

- **Etichetta.** Visualizza tutte le etichette di valore già definite. In questa colonna è possibile aggiungere o modificare etichette.
- **Valore.** Valori univoci per ogni variabile selezionata. L'elenco dei valori univoci è basato sul numero dei casi esaminati. Ad esempio, se nel file di dati sono stati esaminati solo i primi 100 casi, l'elenco conterrà solo i valori univoci presenti in questi casi. Se il file di dati è già stato ordinato in base alla variabile per cui si desidera assegnare etichette di valori, nell'elenco verrà visualizzato un numero di valori univoci notevolmente inferiore a quello dei valori effettivamente presenti nei dati.
- **Conteggio.** Numero di occorrenze di ogni valore nei casi esaminati.
- **Mancante/i.** Valori che rappresentano i dati mancanti. È possibile modificare la designazione di valori mancanti della categoria selezionando la casella di controllo. Un segno di spunta indica che la categoria è definita come categoria di valori mancanti definiti dall'utente. Se una variabile include già un intervallo di valori definiti come valori mancanti definiti dall'utente (ad esempio da 90 a 99), per questa variabile non è possibile aggiungere né eliminare categorie di valori utilizzando Definisci proprietà variabili. Per modificare le categorie di valori mancanti per le variabili contenenti intervalli di valori mancanti, è possibile utilizzare Visualizzazione variabili dell'Editor dei dati. Per ulteriori informazioni, vedere Valori mancanti in Capitolo 5 a pag. 85.
- **Modificato.** Indica che è stata aggiunta o modificata un'etichetta di valore.

Nota: Se nella finestra di dialogo iniziale per il numero dei casi è stato specificato 0, la Griglia etichette di valore sarà inizialmente vuota, eccetto le eventuali etichette di valore già esistenti e/o le categorie di valori mancanti definite per la variabile selezionata. Inoltre, il pulsante Suggerisci per il livello di misurazione risulterà disattivato.

Livello di misurazione. Le etichette di valori sono estremamente utili per le variabili categoriali (nominali e ordinali). Dal momento che alcune procedure utilizzano le variabili di scala e categoriali in modo diverso, è talvolta importante assegnare il corretto livello di misurazione. Tuttavia, a tutte le nuove variabili numeriche viene assegnato un livello di misurazione predefinito. Di conseguenza, è possibile che alcune variabili che sono in realtà categoriali vengano visualizzate come variabili di scala.

Se non si è certi del corretto livello di misurazione da assegnare a una variabile, fare clic su Suggerisci.

Copia proprietà. È possibile copiare etichette di valori e altre proprietà delle variabili da un'altra variabile nella variabile selezionata o dalla variabile selezionata in una o più variabili diverse.

Valori senza etichetta. Per creare automaticamente etichette per valori senza etichetta, fare clic su Etichette automatiche.

Etichette delle variabili e formato di visualizzazione

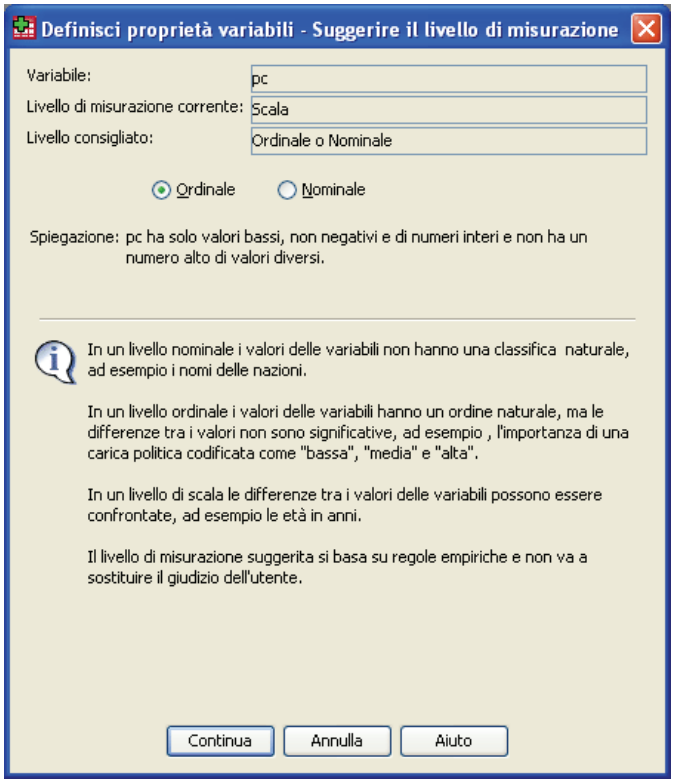
È possibile modificare l'etichetta descrittiva di una variabile e il formato di visualizzazione.

- Non è possibile modificare il tipo fondamentale della variabile (stringa o numerica).
- Per le variabili stringa, è possibile modificare solo l'etichetta e non il formato di visualizzazione.
- Per le variabili numeriche, è possibile modificare il tipo numerico (ad esempio numerico, data, dollaro o valuta personalizzata), larghezza (numero massimo di cifre, inclusi tutti i tipi di separatori di decimali e di indicatori di raggruppamento) e il numero di cifre decimali.
- Per il formato di data numerico, è possibile selezionare un formato di data specifico (ad esempio gg-mm-aaaa, mm/gg/aa e aaaaggg).
- Per il formato numerico personalizzato, è possibile selezionare uno dei cinque formati di valuta personalizzata, da CCA a CCE. Per ulteriori informazioni, vedere Opzioni: Valuta in Capitolo 48 a pag. 528.
- Nella colonna *Valore* viene visualizzato un asterisco se la larghezza specificata è minore della larghezza dei valori esaminati o dei valori visualizzati per le etichette di valori già esistenti o delle categorie di valori mancanti.
- Viene visualizzato un punto (.) se i valori esaminati o i valori visualizzati per le etichette di valori già esistenti o le categorie di valori mancanti non sono valide per il tipo di formato di visualizzazione selezionato. Ad esempio, un valore numerico interno minore di 86.400 non è valido per una variabile nel formato della data.

Assegnazione del livello di misurazione

Quando si fa clic sul pulsante Suggerisci relativo al livello di misurazione nella finestra di dialogo principale Definisci proprietà variabili, la variabile corrente viene valutata in base ai casi esaminati e alle etichette dei valori definite, quindi viene suggerito un livello di misurazione nella finestra di dialogo Suggerisci il livello di misurazione. In Spiegazione viene fornita una breve descrizione dei criteri utilizzati per stabilire il livello di misurazione consigliato.

Figura 7-3
Finestra di dialogo Suggerisci il livello di misurazione


Definisci proprietà variabili - Suggerire il livello di misurazione
Variabile:
pc
Livello di misurazione corrente:
Scala
Livello consigliato:
Ordinale o Nominale
Ordinale
Nominale
Spiegazione: pc ha solo valori bassi, non negativi e di numeri interi e non ha un numero alto di valori diversi.
In un livello nominale i valori delle variabili non hanno una classifica naturale, ad esempio i nomi delle nazioni.
In un livello ordinale i valori delle variabili hanno un ordine naturale, ma le differenze tra i valori non sono significative, ad esempio , l'importanza di una carica politica codificata come "bassa", "media" e "alta".
In un livello di scala le differenze tra i valori delle variabili possono essere confrontate, ad esempio le età in anni.
Il livello di misurazione suggerita si basa su regole empiriche e non va a sostituire il giudizio dell'utente.
Continua
Annulla
Aiuto

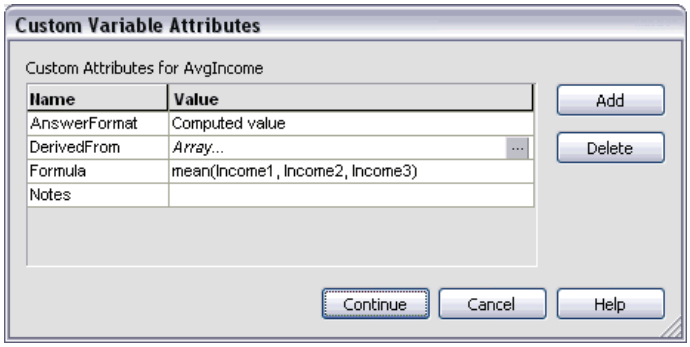
Nota: nella valutazione del livello di misurazione non sono inclusi i valori che rappresentano valori mancanti. Ad esempio, la spiegazione del livello di misurazione può indicare che il suggerimento è in parte basato sull'assenza di valori negativi nella variabile, mentre in effetti è possibile che la variabile contenga valori negativi definiti come valori mancanti.

- Fare clic su Continua per accettare il livello di misurazione suggerito o su Annulla per non accettarlo e mantenere il livello di misurazione stabilito.

Attributi variabili personalizzati

Il pulsante Attributi disponibile in Definisci proprietà variabili apre la finestra di dialogo Attributi variabili. Oltre agli attributi standard delle variabili, come etichette di valore, valori mancanti e livello di misurazione, è possibile creare degli attributi di variabile personalizzati. Analogamente agli attributi standard delle variabili, questi attributi personalizzati vengono salvati in file di dati in formato SPSS Statistics.

Figura 7-4
Attributi variabili personalizzati



Nome. I nomi di attributo devono rispettare le stesse regole applicate ai nomi di variabile. Per ulteriori informazioni, vedere Nomi delle variabili in Capitolo 5 a pag. 79.

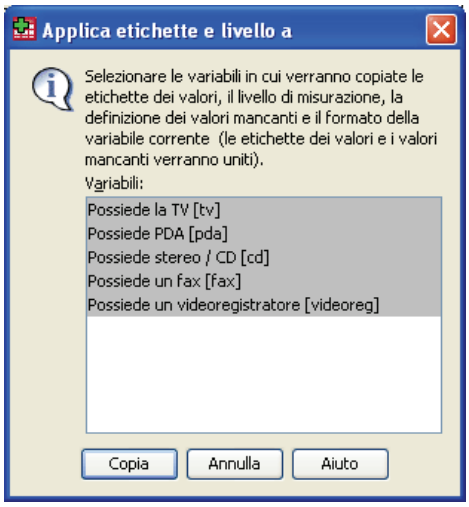
Valore. Il valore assegnato all'attributo per la variabile selezionata.

- I nomi di attributo che iniziano con il simbolo dollaro ($) appartengono ad attributi riservati e non possono essere modificati. È possibile visualizzare il contenuto di un attributo riservato facendo clic sul pulsante nella cella desiderata.
- Il testo *Matrice...* , visualizzato in una cella Valore, indica che si tratta di una **matrice di attributi**, ossia un attributo contenente più valori. Fare clic sul pulsante della cella per visualizzare l'elenco dei valori.

Copia di proprietà delle variabili

La finestra Applica etichette e livello viene visualizzata quando si fa clic su Da un'altra variabile o Ad altre variabili nella finestra di dialogo principale Definisci proprietà variabili. Visualizza tutte le variabili esaminate corrispondenti al tipo di variabile corrente (numerico o stringa). Per le variabili stringa, è necessario che corrisponda anche la larghezza.

Figura 7-5
Finestra di dialogo Applica etichette e livello


Applica etichette e livello a
Selezionare le variabili in cui verranno copiate le etichette dei valori, il livello di misurazione, la definizione dei valori mancanti e il formato della variabile corrente (le etichette dei valori e i valori mancanti verranno uniti).
Variabili:
Possiede la TV [tv]
Possiede PDA [pda]
Possiede stereo / CD [cd]
Possiede un fax [fax]
Possiede un videoregistratore [videoreg]
Copia
Annulla
Aiuto

- ▶ Selezionare una singola variabile da cui copiare le etichette dei valori e altre proprietà della variabile (eccetto l'etichetta della variabile).

  o

- ▶ Selezionare una o più variabili in cui incollare le etichette dei valori e altre proprietà di una variabile.

- ▶ Fare clic su Copia per copiare le etichette dei valori e il livello di misurazione.
  - Le etichette dei valori esistenti e le categorie dei valori mancanti delle variabili di destinazione non verranno sostituite.
  - Le etichette dei valori e le categorie dei valori mancanti relative a valori che non sono ancora stati definiti nelle variabili di destinazione, vengono aggiunte all'insieme di etichette dei valori e categorie di valori mancanti di queste variabili.
  - Il livello di misurazione delle variabili di destinazione viene sempre sostituito.
  - Se la variabile sorgente o la variabile di destinazione include un intervallo di valori mancanti definito, le definizioni dei valori mancanti non vengono copiate.

Insiemi a risposta multipla

Le tabelle personalizzate e il Generatore di grafici supportano un tipo speciale di "variabile" detta **insieme a risposta multipla**. Gli insiemi a risposta multipla non possono essere considerati variabili nell'accezione propria del termine, poiché non è possibile visualizzarli nell'Editor dei dati e non vengono riconosciuti dalle altre procedure. Gli insiemi a risposta multipla utilizzano più variabili per registrare le risposte alle domande in cui il rispondente può fornire più di una risposta. Tali insiemi vengono considerati come variabili categoriali e nella maggior parte dei casi consentono di eseguire le stesse operazioni previste per questo tipo di variabili.

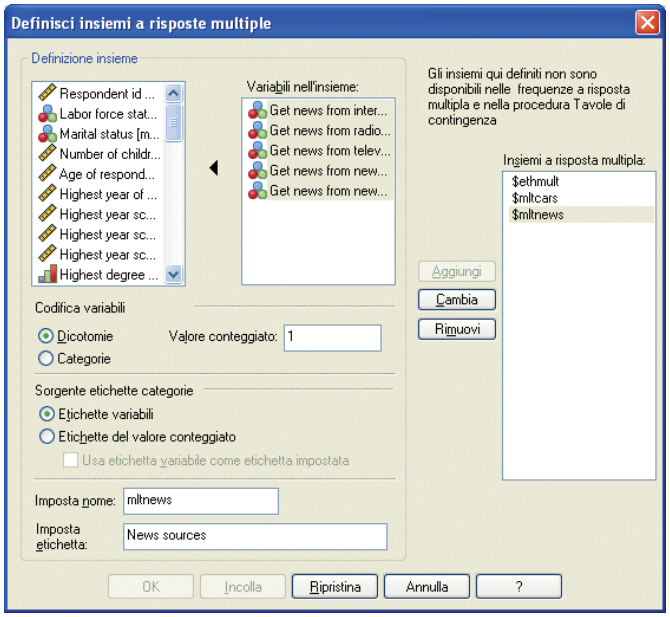
Gli insiemi a risposta multipla vengono creati utilizzando più variabili del file di dati e rappresentano un tipo di insieme speciale all'interno di un file di dati. È possibile definire e salvare più insiemi a risposta multipla in un file di dati in formato SPSS Statistics, ma non è possibile importare o esportare tali insiemi da o in altri formati di file. È possibile copiare gli insiemi a risposta multipla da altri file di dati di SPSS Statistics tramite l'opzione Copia proprietà dati accessibile tramite il menu Dati nella finestra Editor dei dati. Per ulteriori informazioni, vedere Copia delle proprietà dei dati a pag. 117.

Definizione di insiemi a risposta multipla

Per definire gli insiemi a risposta multipla:

- ▶ Dai menu, scegliere:
  Dati
  Definisci insiemi a risposta multipla...

Figura 7-6
Finestra di dialogo Insiemi a risposta multipla



- Selezionare due o più variabili. Se le variabili sono codificate come dicotomie, indicare il valore che si desidera calcolare.
- Immettere un nome univoco per ciascun insieme a risposta multipla. Il nome può contenere fino a 63 byte. All'inizio del nome viene aggiunto automaticamente un segno di dollaro.
- Digitare un'etichetta descrittiva per l'insieme (facoltativo).
- Scegliere Aggiungi per aggiungere l'insieme a risposta multipla all'elenco di insiemi definiti.

Dicotomie

Gli insiemi a dicotomie multiple sono in genere composti da più variabili dicotomiche, ovvero variabili che prevedono solo due possibili valori: sì/no, presente/assente, selezionato/non selezionato. Benché sia possibile che le variabili non siano effettivamente dicotomiche, tutte le variabili dell'insieme sono codificate nello stesso modo e il valore Valore conteggiato rappresenta la condizione positivo/presente/selezionato.

Si consideri, ad esempio, un sondaggio che include la domanda "Quali delle seguenti fonti vengono utilizzate per le informazioni?" per il quale sono disponibili cinque possibili risposte. Il rispondente può specificare più risposte selezionando la casella accanto a ciascuna risposta desiderata. Le cinque risposte sono rappresentate da cinque variabili nel file di dati, codificate con 0 per la risposta *No* (non selezionato) e 1 per la risposta *Sì* (selezionato). Nell'insieme a dicotomie multiple, il valore di Valore conteggiato è 1.

Il file di dati di esempio *survey_sample.sav* contiene già tre insiemi a risposta multipla definiti. *$mltnews* è un insieme a dicotomie multiple.

- Selezionare *$mltnews* nell'elenco Ins. a risposta multipla.

In questo modo vengono visualizzate le variabili e le impostazioni utilizzate per definire l'insieme a risposta multipla specificato.

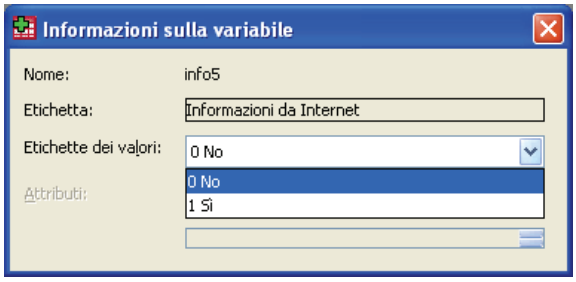
- Nell'elenco Variabili nell'insieme sono visualizzate le cinque variabili utilizzate per creare l'insieme a risposta multipla.
- Nel gruppo Codifica variabili è indicato che le variabili sono dicotomiche.
- Il valore di Valore conteggiato è 1.

▶ Selezionare una delle variabili nell'elenco Variabili nell'insieme.

▶ Fare clic con il pulsante destro del mouse sulla variabile e scegliere Informazioni sulla variabile dal menu di scelta rapida.

▶ Nella finestra Informazioni sulla variabile fare clic sulla freccia dell'elenco a discesa Etichette di valore per visualizzare l'elenco completo delle etichette di valore definite.

Figura 7-7
Informazioni sulla variabile per la variabile sorgente delle dicotomie multiple



Le etichette di valore indicano che la variabile è dicotomica, con i valori 0 e 1 che rappresentano rispettivamente *No* e *Sì*. Tutte e cinque le variabili dell'elenco sono codificate nello stesso modo e il valore 1 (codice per *Sì*) è il valore conteggiato per l'insieme a dicotomie multiple.

Categorie

Un insieme a categorie multiple è composto da più variabili codificate nello stesso modo e spesso prevede numerose categorie possibili di risposta. Si consideri, ad esempio, un sondaggio che prevede la seguente specifica: "Indicare fino a tre nazionalità che meglio descrivono la propria identità etnica". Le risposte possibili sono centinaia, ma per motivi di codifica l'elenco è limitato a 40 nazionalità e include una categoria "Altro" che consente di specificare voci diverse. Nel file di dati, le tre scelte possibili rappresentano tre variabili, ciascuna delle quali include 41 categorie (40 nazionalità codificate e la categoria "Altro").

Nel file di dati di esempio, *$ethmult* e *$mltcars* sono insiemi a categorie multiple.

Sorgente etichetta categoria

Per le dicotomie multiple è possibile controllare la modalità di etichettatura degli insiemi.

- **Etichette delle variabili.** Le etichette di variabili definite (o nomi delle variabili per le variabili prive di etichette di variabili definite) vengono utilizzate come etichette di categoria dell'insieme. Ad esempio, se tutte le variabili dell'insieme possiedono la stessa etichetta di

valore (o nessuna etichetta di valore definita) per il valore conteggiato (ad esempio, *Si*), è necessario utilizzare le etichette di variabili come etichette di categoria dell'insieme.

- **Etichette dei valori conteggiati.** Le etichette di valore definite dei valori conteggiati vengono utilizzate come etichette di categoria dell'insieme. Selezionare questa opzione solo se tutte le variabili possiedono un'etichetta di valore definita per il valore conteggiato e tale etichetta è diversa per ogni variabile.
- **Utilizzare l'etichetta di variabile come etichetta dell'insieme.** Se si seleziona Etichette dei valori conteggiati, è inoltre possibile utilizzare l'etichetta di variabile per la prima variabile dell'insieme che possiede un'etichetta di variabile definita come etichetta dell'insieme. Se nessuna delle variabili dell'insieme possiede un'etichetta di variabile definita, il nome della prima variabile dell'insieme viene utilizzata come etichetta dell'insieme stesso.

Copia delle proprietà dei dati

La Copia proprietà dei dati guidata consente di utilizzare un file di dati di SPSS Statistics esterno come modello per definire le proprietà di file e variabili nell'insieme di dati attivo. È possibile anche utilizzare le variabili nell'insieme di dati attivo come modelli per altre variabili dello stesso insieme di dati. È possibile:

- Copiare le proprietà del file selezionate da un file di dati esterno o da un insieme di dati aperto nell'insieme di dati attivo. Le proprietà dei file includono documenti, etichette dei file, insiemi a risposta multipla, insiemi di variabili e ponderazione.
- Copiare le proprietà della variabile selezionata da un file di dati esterno o da un insieme di dati aperto verificando la corrispondenza con le variabili nell'insieme di dati attivo. Le proprietà delle variabili includono etichette dei valori, valori mancanti, livello di misurazione, etichette delle variabili, formati di stampa e di scrittura, allineamento e larghezza delle colonne (nell'Editor dei dati).
- Copiare le proprietà delle variabili selezionate di una variabile contenuta in un file di dati esterno, un insieme di dati aperto o nell'insieme di dati attivo in più variabili dell'insieme di dati attivo.
- Creare nuove variabili nell'insieme di dati attivo in base alle variabili selezionate in un file di dati esterno o un insieme di dati aperto.

Quando si copiano le proprietà dei dati, è necessario seguire le seguenti regole:

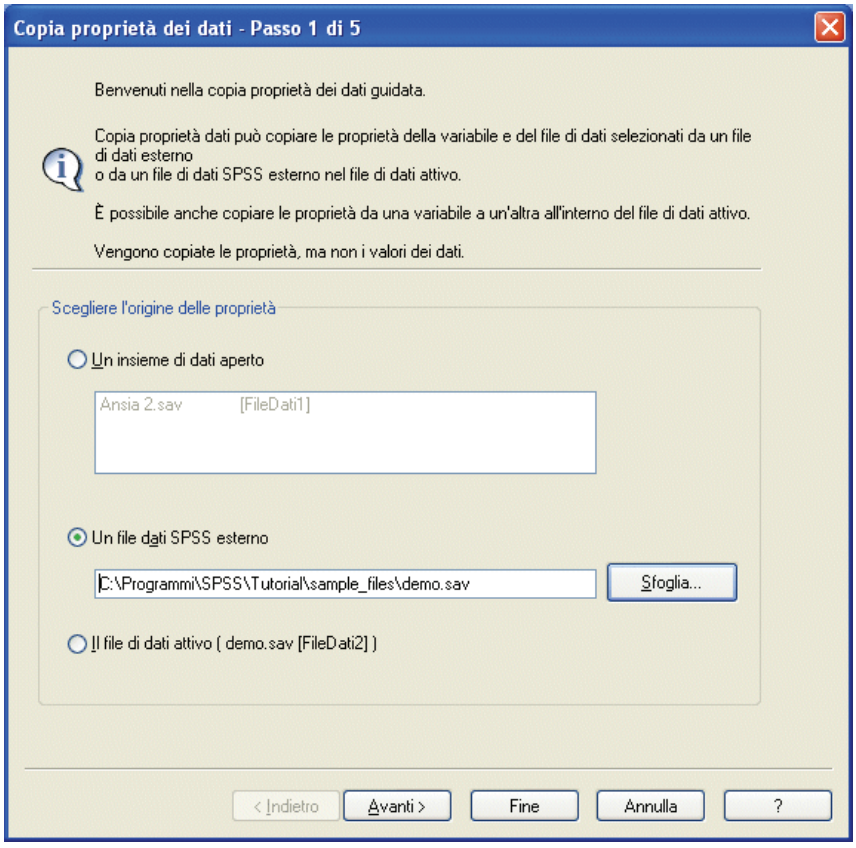
- Se come file di dati sorgente si utilizza un file di dati esterno, questo deve essere in formato SPSS Statistics.
- Se come file di dati sorgente si utilizza l'insieme di dati attivo, questo deve includere almeno una variabile. Non è possibile usare insiemi di dati attivi completamente vuoti come file di dati sorgente.
- Le proprietà non definite (vuote) nell'insieme di dati sorgente non sovrascrivono le proprietà definite nell'insieme di dati attivo.
- Le proprietà delle variabili vengono copiate dalla variabile sorgente solo nelle variabili di destinazione di tipo corrispondente, stringa (alfanumerica) o numerica (numerica, data e valuta).

Nota: La funzione Copia proprietà dei dati sostituisce la funzione Applica dizionario dati, precedentemente disponibile nel menu File.

Per copiare le proprietà dei dati

- Dai menu nella finestra dell'Editor dei dati, scegliere:
  Dati
  Copia proprietà dei dati...

Figura 7-8
Copia proprietà dei dati guidata: Passo 1

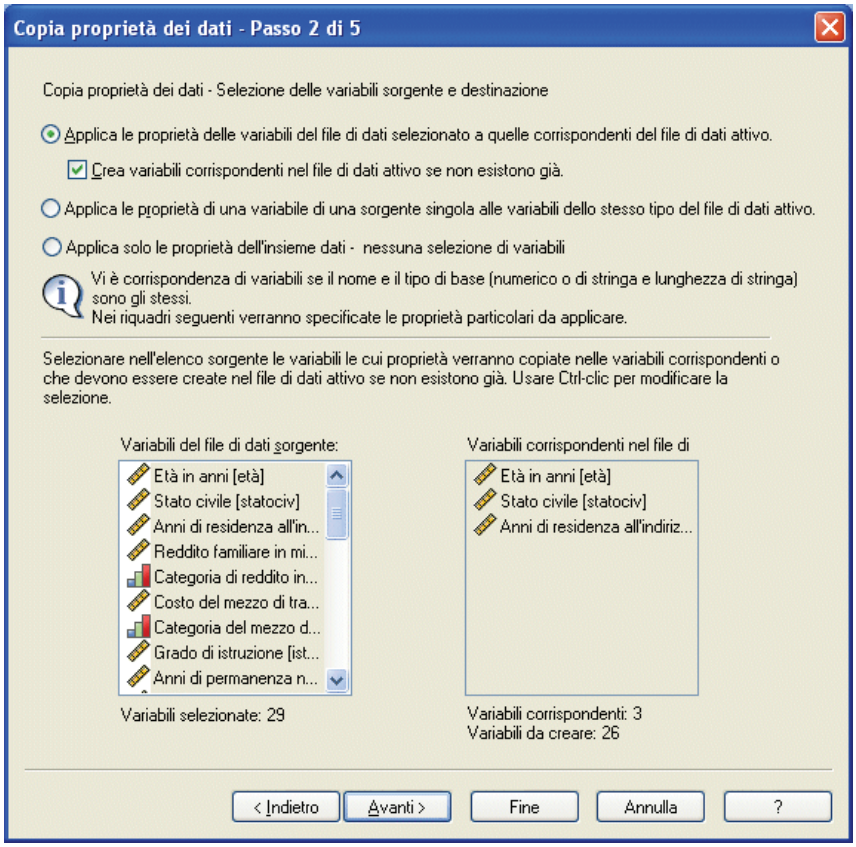

Copia proprietà dei dati - Passo 1 di 5
Benvenuti nella copia proprietà dei dati guidata.
Copia proprietà dati può copiare le proprietà della variabile e del file di dati selezionati da un file di dati esterno
o da un file di dati SPSS esterno nel file di dati attivo.
È possibile anche copiare le proprietà da una variabile a un'altra all'interno del file di dati attivo.
Vengono copiate le proprietà, ma non i valori dei dati.
Scegliere l'origine delle proprietà
Un insieme di dati aperto
Ansia 2.sav [FileDati1]
Un file dati SPSS esterno
C:\Programmi\SPSS\Tutorial\sample_files\demo.sav
Sfoglia...
Il file di dati attivo (demo.sav [FileDati2])
< Indietro
Avanti >
Fine
Annulla
?

- Selezionare il file di dati e/o le proprietà della variabile da copiare. Il file può essere un insieme di dati aperto, un file di dati esterno in formato SPSS Statistics o l'insieme di dati attivo.
- Eseguire le istruzioni visualizzate nella procedura guidata per la copia delle proprietà dei dati.

Selezione delle variabili sorgente e destinazione

In questo passaggio è possibile specificare le variabili sorgente che includono le proprietà delle variabili che si desidera copiare e le variabili di destinazione che riceveranno le proprietà in questione.

Figura 7-9
Copia proprietà dei dati guidata: Passo 2


Copia proprietà dei dati - Passo 2 di 5
Copia proprietà dei dati - Selezione delle variabili sorgente e destinazione
Applica le proprietà delle variabili del file di dati selezionato a quelle corrispondenti del file di dati attivo.
Crea variabili corrispondenti nel file di dati attivo se non esistono già.
Applica le proprietà di una variabile di una sorgente singola alle variabili dello stesso tipo del file di dati attivo.
Applica solo le proprietà dell'insieme dati - nessuna selezione di variabili
Vi è corrispondenza di variabili se il nome e il tipo di base (numerico o di stringa e lunghezza di stringa) sono gli stessi.
Nei riquadri seguenti verranno specificate le proprietà particolari da applicare.
Selezionare nell'elenco sorgente le variabili le cui proprietà verranno copiate nelle variabili corrispondenti o che devono essere create nel file di dati attivo se non esistono già. Usare Ctrl-clic per modificare la selezione.
Variabili del file di dati sorgente:
Età in anni [età]
Stato civile [statociv]
Anni di residenza all'in...
Reddito familiare in mi...
Categoria di reddito in...
Costo del mezzo di tra...
Categoria del mezzo d...
Grado di istruzione [ist...
Anni di permanenza n...
Variabili selezionate: 29
Variabili corrispondenti nel file di
Età in anni [età]
Stato civile [statociv]
Anni di residenza all'indiriz...
Variabili corrispondenti: 3
Variabili da creare: 26
< Indietro
Avanti >
Fine
Annulla
?

Applica le proprietà delle variabili dell'insieme di dati selezionato a quelle corrispondenti dell'insieme di dati attivo. Le proprietà delle variabili vengono copiate da una o più variabili sorgente selezionate nelle variabili corrispondenti dell'insieme di dati attivo. Le variabili "corrispondono" quando condividono lo stesso nome e tipo di variabile (stringa o numerica). Per le variabili stringa, deve corrispondere anche la lunghezza definita. Per impostazione predefinita, nei due elenchi di variabili vengono visualizzate solo le variabili corrispondenti.

- **Crea variabili corrispondenti nell'insieme di dati attivo se non esistono già.** L'elenco sorgente verrà aggiornato in modo da includere tutte le variabili del file di dati sorgente. Se si selezionano variabili sorgente che non esistono nell'insieme di dati attivo (in base al nome della variabile), nell'insieme di dati verranno create nuove variabili con nomi e proprietà corrispondenti a quelli delle variabili del file di dati sorgente.

Se l'insieme di dati attivo non include variabili (nuovo file di dati vuoto), vengono visualizzate tutte le variabili del file di dati sorgente e nell'insieme di dati attivo vengono create nuove variabili in base alle variabili sorgente selezionate.

Applica le proprietà di una variabile di una sorgente singola alle variabili dello stesso tipo dell'insieme di dati attivo. È possibile applicare le proprietà delle variabili di una singola variabile selezionata nell'elenco sorgente a una o più variabili selezionate nell'elenco dell'insieme di dati attivo. Nell'elenco dell'insieme di dati attivo vengono visualizzate solo variabili dello stesso tipo

(numeriche o stringa) della variabile selezionata nell'elenco sorgente. Per le variabili stringa, vengono visualizzate solo le stringhe che presentano la stessa lunghezza definita della variabile sorgente. Questa opzione non è disponibile se l'insieme di dati attivo non contiene variabili.

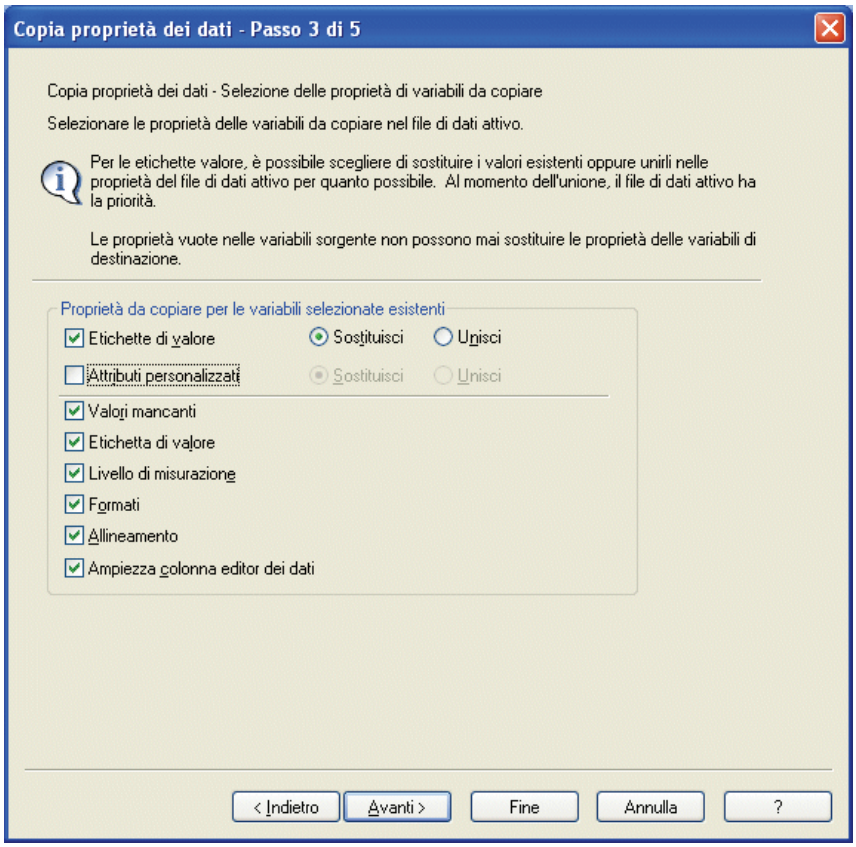
Nota: non è possibile creare nuove variabili nell'insieme di dati attivo con questa opzione.

Applicare solo variabili dell'insieme di dati, senza selezione di variabili. All'insieme di dati attivo verranno applicate solo le proprietà dei file (ad esempio, documenti, etichette di file, peso). Non verranno applicate proprietà delle variabili. Questa opzione non è disponibile se l'insieme di dati attivo è anche la sorgente del file di dati stesso.

Selezione delle proprietà delle variabili da copiare

È possibile copiare le proprietà delle variabili selezionate dalle variabili sorgente alle variabili di destinazione. Le proprietà non definite (vuote) delle variabili sorgente non sovrascrivono le proprietà definite delle variabili di destinazione.

Figura 7-10
Copia proprietà dei dati guidata: Passo 3


Copia proprietà dei dati - Passo 3 di 5
Copia proprietà dei dati - Selezione delle proprietà di variabili da copiare
Selezionare le proprietà delle variabili da copiare nel file di dati attivo.
Per le etichette valore, è possibile scegliere di sostituire i valori esistenti oppure unirli nelle proprietà del file di dati attivo per quanto possibile. Al momento dell'unione, il file di dati attivo ha la priorità.
Le proprietà vuote nelle variabili sorgente non possono mai sostituire le proprietà delle variabili di destinazione.
Proprietà da copiare per le variabili selezionate esistenti
Etichette di valore
Sostituisci
Unisci
Attributi personalizzati
Sostituisci
Unisci
Valori mancanti
Etichetta di valore
Livello di misurazione
Formati
Allineamento
Ampiezza colonna editor dei dati
< Indietro
Avanti >
Fine
Annulla
?

Etichette dei valori. Le etichette dei valori sono etichette descrittive associate ai valori dei dati. Le etichette dei valori vengono utilizzate frequentemente per rappresentare con valori di dati numerici categorie non numeriche (ad esempio, i codici 1 e 2 per *Maschile* e *Femminile*). È possibile sostituire o unire le etichette dei valori nelle variabili di destinazione.

- Il comando Sostituisci consente di eliminare tutte le etichette dei valori definite per la variabile di destinazione e di sostituirle con le etichette dei valori definite per la variabile sorgente.
- Il comando Unisci consente di unire tutte le etichette dei valori definite per la variabile sorgente alle etichette dei valori definite per la variabile di destinazione. Se allo stesso valore corrisponde un'etichetta di valore definita nelle variabili sia sorgente sia di destinazione, l'etichetta del valore nella variabile di destinazione rimarrà invariata.

Attributi personalizzati. Attributi delle variabili definite dall'utente. Per ulteriori informazioni, vedere Attributi variabili personalizzati in Capitolo 5 a pag. 87.

- Il comando Sostituisci consente di eliminare gli attributi personalizzati per la variabile di destinazione e di sostituirli con gli attributi definiti nella variabile sorgente.
- Il comando Unisci consente di unire gli attributi definiti per la variabile sorgente agli attributi definiti esistenti definiti per la variabile di destinazione.

Valori mancanti. I valori mancanti sono valori utilizzati per rappresentare dati mancanti (ad esempio, 98 per *Non so* e 99 per *Non applicabile*). In genere a questi valori sono associate etichette definite che indicano a cosa si riferiscono i codici dei valori mancanti. Tutti i valori mancanti definiti per la variabile di destinazione vengono eliminati e sostituiti con i valori mancanti definiti per la variabile sorgente.

Etichette delle variabili. Le etichette delle variabili descrittive possono contenere spazi e caratteri riservati non consentiti nei nomi delle variabili. Se si copiano le proprietà delle variabili da una singola variabile sorgente in più variabili di destinazione, è necessario valutare accuratamente l'opportunità di utilizzare questa opzione.

Livello di misurazione. Il livello di misurazione può essere nominale, ordinale o di scala. Nelle procedure in cui è rilevante la distinzione tra i diversi livelli di misurazione, i livelli nominale e ordinale vengono entrambi considerati **categoriali**.

Formati. Per le variabili numeriche, questa opzione controlla il tipo di numeri (ad esempio numerico, data o valuta), la larghezza (numero totale di caratteri visualizzati, inclusi i caratteri di testa e di coda e l'indicatore decimale) e il numero di cifre decimali visualizzate. Questa opzione viene ignorata per le variabili stringa.

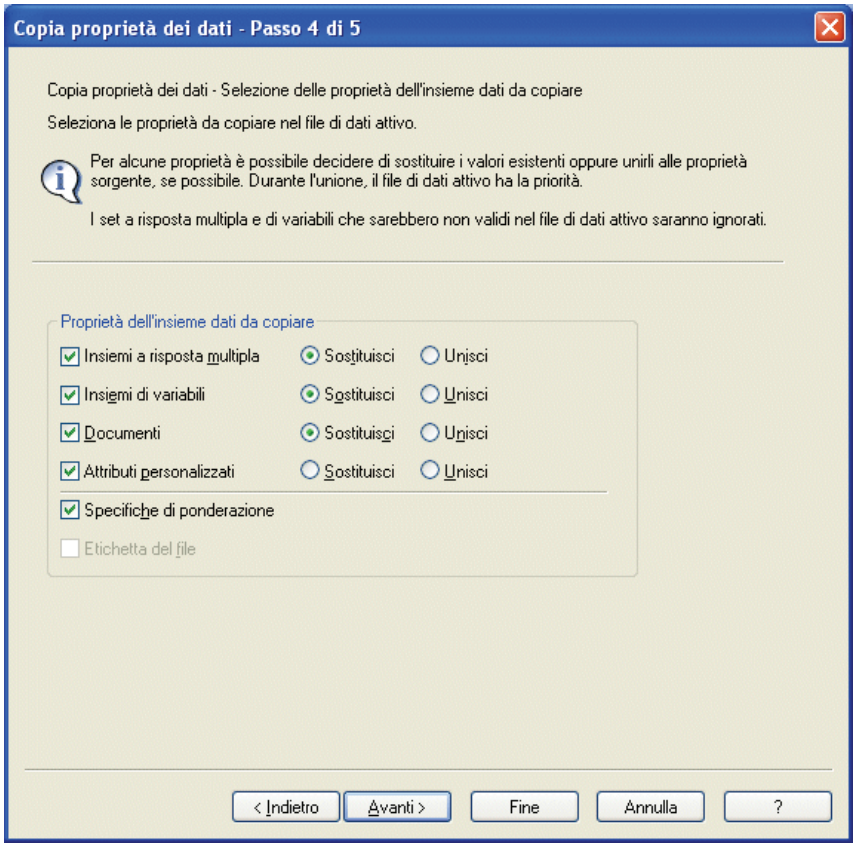
Allineamento. Questa opzione influisce solo sull'allineamento (a sinistra, a destra, al centro) nella Visualizzazione dati dell'Editor dei dati.

Larghezza delle colonne nell'Editor dei dati. Questa opzione influisce solo sulla larghezza delle colonne nella Visualizzazione dati dell'Editor dei dati.

Copia delle proprietà dell'insieme di dati (file)

È possibile applicare all'insieme di dati attivo le proprietà globali dell'insieme di dati selezionate nel file di dati sorgente. Questa opzione non è disponibile se l'insieme di dati attivo è anche la sorgente del file di dati stesso.

Figura 7-11
Copia proprietà dei dati guidata: Passo 4


Copia proprietà dei dati - Passo 4 di 5
Copia proprietà dei dati - Selezione delle proprietà dell'insieme dati da copiare
Seleziona le proprietà da copiare nel file di dati attivo.
Per alcune proprietà è possibile decidere di sostituire i valori esistenti oppure unirli alle proprietà sorgente, se possibile. Durante l'unione, il file di dati attivo ha la priorità.
I set a risposta multipla e di variabili che sarebbero non validi nel file di dati attivo saranno ignorati.
Proprietà dell'insieme dati da copiare
Insiemi a risposta multipla
Sostituisci
Unisci
Insiemi di variabili
Sostituisci
Unisci
Documenti
Sostituisci
Unisci
Attributi personalizzati
Sostituisci
Unisci
Specifiche di ponderazione
Etichetta del file
< Indietro
Avanti >
Fine
Annulla
?

Insiemi a risposta multipla. Applica all'insieme di dati attivo le definizioni dell'insieme a risposta multipla del file di dati sorgente.

- Gli insiemi a risposta multipla del file di dati sorgente che includono variabili che non esistono nell'insieme di dati attivo vengono ignorati, a meno che non vengano create le variabili necessarie in base alle indicazioni fornite nel passo 2 (Selezione delle variabili sorgente e di destinazione) della Copia proprietà dei dati guidata.
- Il comando Sostituisci consente di eliminare tutti gli insiemi a risposta multipla inclusi nell'insieme di dati attivo e di sostituirli con gli insiemi a risposta multipla del file di dati sorgente.
- Il comando Unisci consente di unire gli insiemi a risposta multipla del file di dati di origine al gruppo di insiemi a risposta multipla dell'insieme di dati attivo. Se entrambi i file includono un insieme con lo stesso nome, l'insieme di dati attivo rimarrà invariato.

Insiemi di variabili. Gli insiemi di variabili vengono utilizzati per controllare l'elenco delle variabili visualizzate nelle finestre di dialogo. Gli insiemi di variabili vengono definiti scegliendo Definisci insiemi dal menu Strumenti.

- Gli insiemi del file di dati sorgente che includono variabili che non esistono nell'insieme di dati attivo verranno ignorati, a meno che tali variabili non vengano create in base alle indicazioni del passo 2 (Selezione di variabili sorgente e di destinazione) della Copia proprietà dei dati guidata.
- Il comando Sostituisci consente di eliminare gli insiemi di variabili esistenti nell'insieme di dati attivo e di sostituirli con gli insiemi di variabili del file di dati sorgente.
- Il comando Unisci consente di aggiungere insiemi di variabili del file di dati sorgente al gruppo di insiemi di variabili dell'insieme di dati attivo. Se entrambi i file includono un insieme con lo stesso nome, l'insieme di dati attivo rimarrà invariato.

Documenti. Note aggiunte al file di dati mediante il comando `DOCUMENT`.

- Il comando Sostituisci consente di eliminare i documenti esistenti nell'insieme di dati attivo e di sostituirli con i documenti del file di dati sorgente.
- Il comando Unisci consente di combinare i documenti dei file di dati sorgente e dell'insieme di dati attivo. All'insieme di dati attivo vengono aggiunti i documenti univoci del file di dati sorgente che non esistono al suo interno. Tutti i documenti vengono quindi ordinati in base alla data.

Attributi personalizzati. Attributi dei file di dati personalizzati, generalmente creati dal comando `DATAFILE ATTRIBUTE` nella sintassi dei comandi.

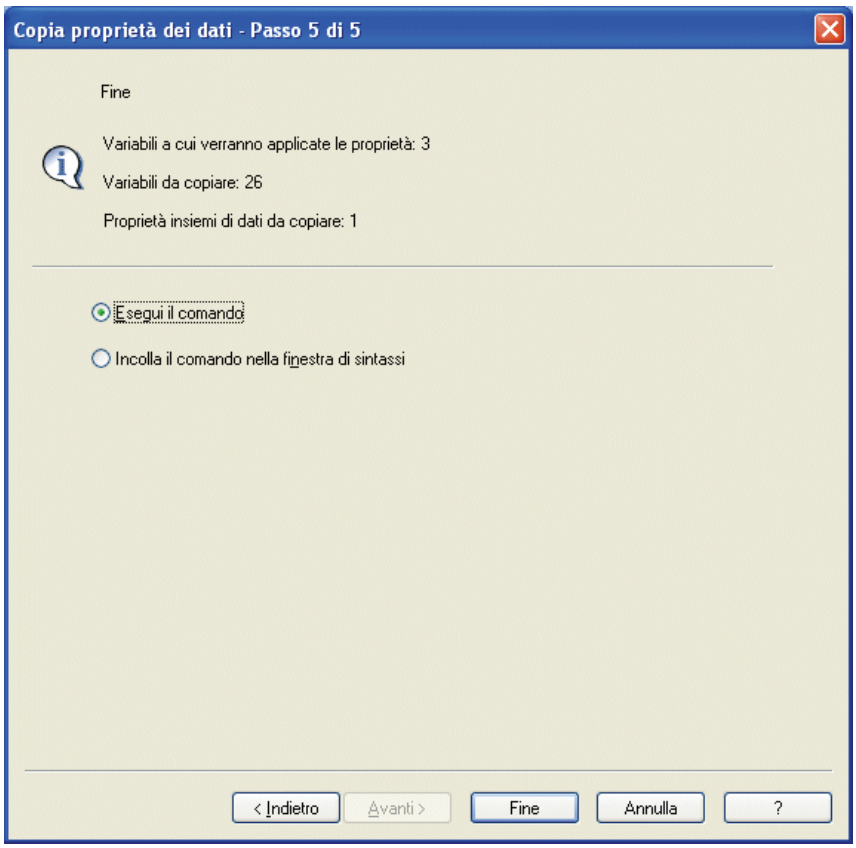
- Sostituisci elimina tutti gli attributi dei file di dati personalizzati esistenti nell'insieme di dati attivo, sostituendoli con quelli del file di dati sorgente.
- Il comando Unisci consente di combinare gli attributi del file di dati sorgente e del'insieme di dati attivo. All'insieme di dati attivo vengono aggiunti i nomi degli attributi univoci del file di dati sorgente che non esistono al suo interno. Se il nome dell'attributo è già presente in entrambi i file di dati, l'attributo indicato nell'insieme di dati attivo rimane invariato.

Specifiche di ponderazione. Esegue la ponderazione dei casi in base alla variabile di ponderazione corrente del file di dati sorgente, se esiste una variabile corrispondente nell'insieme di dati attivo. Questa operazione ha la priorità rispetto a qualsiasi processo di ponderazione in atto nell'insieme di dati attivo.

Etichetta del file. Etichetta descrittiva applicata a un file di dati mediante il comando `FILE LABEL`.

Risultati

Figura 7-12
Copia proprietà dei dati guidata: Passo 5


Copia proprietà dei dati - Passo 5 di 5
Fine
Variabili a cui verranno applicate le proprietà: 3
Variabili da copiare: 26
Proprietà insiemi di dati da copiare: 1
Esegui il comando
Incolla il comando nella finestra di sintassi
< Indietro
Avanti >
Fine
Annulla
?

Nell'ultimo passo della Copia proprietà dei dati guidata vengono fornite informazioni sul numero di variabili per cui verranno copiate proprietà dal file di dati sorgente, sul numero di variabili che verranno create e sul numero di proprietà dell'insieme di dati (file) che verranno copiate.

È inoltre possibile incollare la sintassi dei comandi generata in una finestra di sintassi e salvarla per un successivo utilizzo.

Identificazione di casi duplicati

La presenza di casi duplicati nei dati può essere dovuta a svariati motivi, ad esempio:

- Errori nell'inserimento dei dati, che comportano l'inserimento accidentale dello stesso caso più volte.
- Più casi che condividono lo stesso ID principale ma con ID secondari diversi, ad esempio i membri di una famiglia che risiedono nella stessa abitazione.
- Più casi che rappresentano lo stesso caso, ma per i quali sono stati impostati valori diversi per le variabili diverse dalle variabili di identificazione del caso, ad esempio nel caso di più acquisti effettuati dalla stessa persona o società in relazione a prodotti diversi o in momenti diversi.

L'opzione Identifica i casi duplicati consente di definire i *duplicati* nel modo desiderato e di gestire la determinazione automatica dei casi principali e dei casi duplicati.

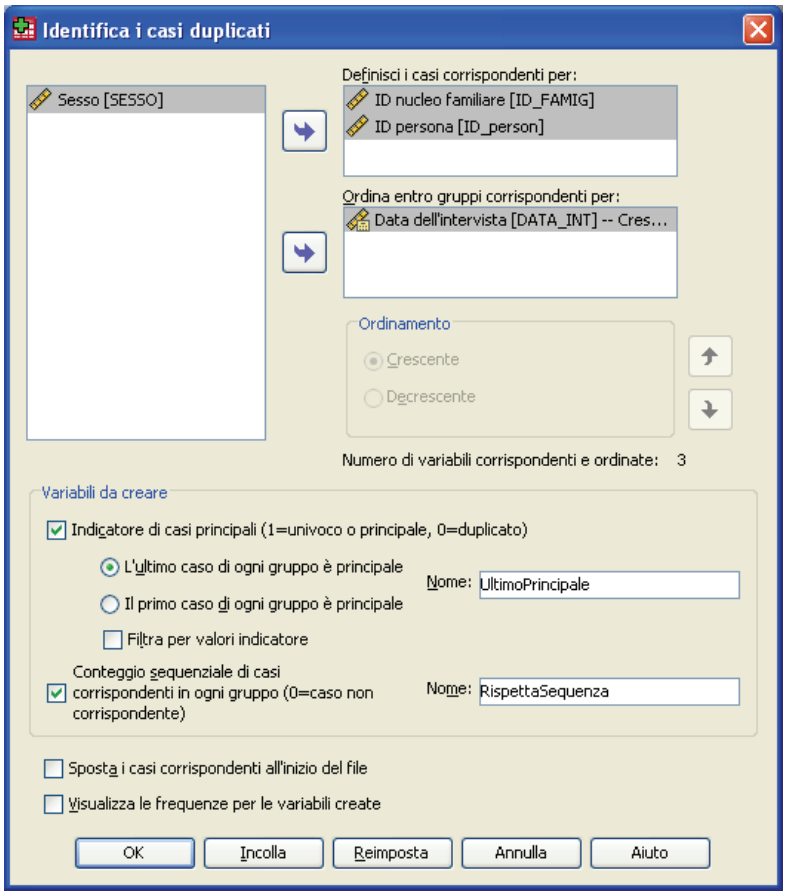
Per identificare e contrassegnare i casi duplicati:

- Dai menu, scegliere:
  Dati
  Identifica casi duplicati...
- Selezionare una o più variabili che identificano i casi corrispondenti.
- Selezionare una o più opzioni del gruppo Variabili da creare.

Se lo si desidera, è possibile:

- Selezionare una o più variabili per ordinare i casi entro i gruppi definiti dalle variabili dei casi corrispondenti selezionate. L'ordinamento definito da tali variabili determina il primo e l'ultimo caso di ogni gruppo. In caso contrario, verrà utilizzato l'ordinamento del file originale.
- Filtrare automaticamente i casi duplicati in modo tale che non vengano inseriti nei rapporti, nei grafici o nei calcoli delle statistiche.

Figura 7-13
Finestra di dialogo Identifica i casi duplicati



Definisci i casi corrispondenti per. I casi vengono considerati come duplicati se i relativi valori per *tutte* le variabili selezionate risultano corrispondenti. Se si desidera identificare solo i casi corrispondenti al 100%, selezionare tutte le variabili.

Ordina entro gruppi corrispondenti per. I casi vengono ordinati automaticamente in base alle variabili che definiscono i casi corrispondenti. È possibile selezionare variabili di ordinamento aggiuntive in base alle quali verrà determinato l'ordine sequenziale dei casi in ogni gruppo corrispondente.

- Per ogni variabile di ordinamento è possibile eseguire l'ordinamento crescente o decrescente.
- Se si selezionano più variabili di ordinamento, i casi verranno ordinati in base a ciascuna variabile all'interno delle categorie della variabile precedente nell'elenco. Se, ad esempio, si seleziona *data* come prima variabile di ordinamento e *quantità* come seconda variabile di ordinamento, i casi verranno classificati per quantità all'interno di ciascuna categoria di data.
- Per modificare l'ordine delle variabili, utilizzare le frecce in su e in giù a destra dell'elenco.
- L'ordinamento determina il primo e l'ultimo caso all'interno di ogni gruppo corrispondente, che a sua volta determina il valore della variabile indicatore principale facoltativa. Se, ad esempio, si desidera filtrare tutti i casi tranne il più recente di ogni gruppo corrispondente, è possibile ordinare i casi all'interno del gruppo in ordine crescente in base a una variabile di data. In questo modo, la data più recente risulterà essere l'ultima data nel gruppo.

Indicatore di casi principali. Crea una variabile con il valore 1 per tutti i casi univoci e il caso identificato come il caso principale in ogni gruppo di casi corrispondenti e con un valore 0 per i duplicati non principali in ogni gruppo.

- Il caso principale può essere l'ultimo o il primo caso di ogni gruppo corrispondente, in base all'ordinamento all'interno del gruppo corrispondente. Se non si specificano variabili di ordinamento, l'ordine dei casi entro ogni gruppo sarà determinato dall'ordinamento del file originale.
- È possibile utilizzare la variabile indicatore come una **variabile filtro** per escludere i duplicati non principali dai rapporti e dalle analisi senza eliminare i casi dal file di dati.

Conteggio sequenziale di caso corrisp. in ogni gruppo. Crea una variabile con un valore sequenziale compreso tra 1 e *n* per i casi di ogni gruppo corrispondente. La sequenza si basa sull'ordine corrente dei casi in ogni gruppo, che è rappresentato dall'ordinamento del file originale oppure dall'ordine determinato dalle variabili di ordinamento specificate.

Sposta i casi corrispondenti all'inizio del file. Ordina il file dati in modo tale che tutti i gruppi di casi corrispondenti si trovino all'inizio del file, semplificando il controllo visuale dei casi corrispondenti nell'Editor dei dati.

Visualizza le frequenze per la variabili create. Tabelle di frequenza contenenti i conteggi relativi a ogni valore delle variabili create. Ad esempio, per la variabile indicatore principale, nella tabella verrà visualizzato il numero di casi con il valore 0 per la variabile, a indicare il numero di duplicati, e il numero di casi con il valore 1, a indicare il numero di casi univoci e principali.

Valori mancanti. Per le variabili numeriche, il valore mancante di sistema viene considerato come qualsiasi altro valore.—I casi con un valore mancante di sistema per una variabile identificatrice vengono considerati come casi che hanno valori corrispondenti per la variabile specificata. Per le

variabili stringa, i casi senza valori per una variabile identificatrice vengono considerati come casi che hanno valori corrispondenti per la variabile.

Categorizzazione visuale

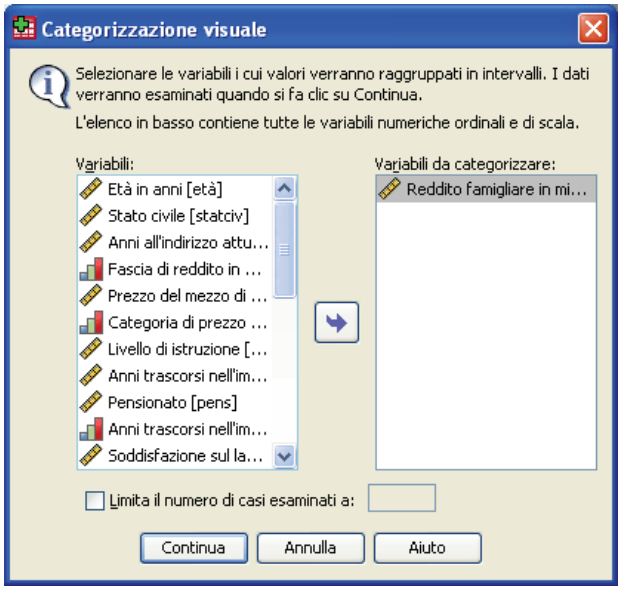
L'utilità Categorizzazione visuale è progettata per semplificare il processo di creazione di nuove variabili basate sul raggruppamento di valori contigui di variabili esistenti in un numero limitato di categorie distinte. È possibile utilizzare questo metodo per eseguire le operazioni seguenti:

- Creare variabili categoriali da variabili di scala continue. Ad esempio, è possibile utilizzare una variabile continua relativa al reddito per creare una nuova variabile categoriale contenente intervalli di reddito.
- Comprimere un elevato numero di categorie ordinali in un insieme più piccolo di categorie. Ad esempio, è possibile comprimere una scala di stime di nove categorie in tre categorie che rappresentino i valori bassi, medi ed elevati.

Nel primo passaggio:

▶ Selezionare la scala numerica e/o le variabili ordinali per cui si desidera creare le nuove variabili categoriali (categorizzate).

Figura 7-14
Finestra di dialogo iniziale per la selezione delle variabili da categorizzare


Categorizzazione visuale
Selezionare le variabili i cui valori verranno raggruppati in intervalli. I dati verranno esaminati quando si fa clic su Continua.
L'elenco in basso contiene tutte le variabili numeriche ordinali e di scala.
Variabili:
Età in anni [età]
Stato civile [statciv]
Anni all'indirizzo attu...
Fascia di reddito in ...
Prezzo del mezzo di ...
Categoria di prezzo ...
Livello di istruzione [...
Anni trascorsi nell'im...
Pensionato [pens]
Anni trascorsi nell'im...
Soddisfazione sul la...
Variabili da categorizzare:
Reddito famigliare in mi...
Limita il numero di casi esaminati a:
Continua
Annulla
Aiuto

È inoltre possibile limitare il numero di casi da esaminare. Per file di dati con un elevato numero di casi, la limitazione del numero di casi esaminati consente di risparmiare tempo. Tuttavia è consigliabile evitarla, se possibile, perchè influenza la distribuzione dei valori utilizzati per calcoli successivi in Categorizzazione visuale.

Nota: le variabili stringa e le variabili nominali numeriche non vengono visualizzate nell'elenco di variabili sorgente. L'utilità Categorizzazione visuale richiede variabili numeriche, misurate su una scala o un livello ordinale, in quanto presume che i valori dei dati rappresentino un tipo di

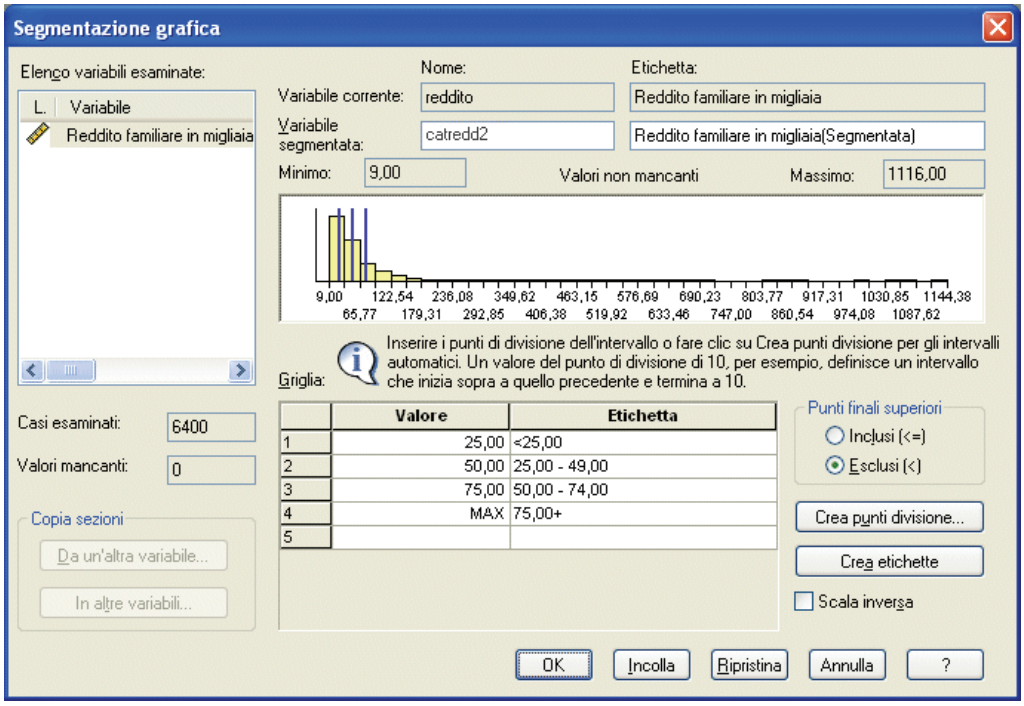
ordine logico che può essere utilizzato per raggruppare valori in modo significativo. È possibile modificare il livello di misurazione definito per una variabile nella Visualizzazione variabili dell'Editor dei dati. Per ulteriori informazioni, vedere Livello di misurazione delle variabili in Capitolo 5 a pag. 80.

Per categorizzare le variabili

- Dai menu nella finestra dell'Editor dei dati, scegliere:
  Trasforma
  Categorizzazione visuale...
- Selezionare la scala numerica e/o le variabili ordinali per cui si desidera creare le nuove variabili categoriali (categorizzate).
- Selezionare una variabile dall'Elenco variabili esaminate.
- Specificare un nome per la nuova variabile categorizzata. I nomi delle variabili devono essere univoci e rispettare le regole di denominazione delle variabili. Per ulteriori informazioni, vedere Nomi delle variabili in Capitolo 5 a pag. 79.
- Definire il criterio di categorizzazione per la nuova variabile. Per ulteriori informazioni, vedere Categorizzazione di variabili a pag. 128.
- Fare clic su OK.

Categorizzazione di variabili

Figura 7-15
Finestra di dialogo principale Categorizzazione visuale



La finestra di dialogo principale Categorizzazione visuale include le seguenti informazioni sulle variabili esaminate:

Elenco variabili esaminate. Visualizza le variabili selezionate nella finestra di dialogo iniziale. È possibile ordinare l'elenco in base al livello di misurazione (di scala o ordinale) oppure in base al nome o all'etichetta delle variabili facendo clic sulle intestazioni delle relative colonne.

Casi esaminati. Indica il numero di casi esaminati. Tutti i casi esaminati privi di valori mancanti di sistema o definiti dall'utente per la variabile selezionata vengono utilizzati per generare la distribuzione dei valori utilizzati per i calcoli in Categorizzazione visuale, incluso l'istogramma visualizzato nella finestra di dialogo principale e i punti di divisione basati su percentili o unità di deviazione standard.

Valori mancanti. Indica il numero di casi esaminati con valori mancanti di sistema o valori mancanti definiti dall'utente. I valori mancanti vengono esclusi da tutte le categorie raggruppate. Per ulteriori informazioni, vedere Valori mancanti definiti dall'utente in Categorizzazione visuale a pag. 134.

Variabile corrente. Il nome e l'eventuale etichetta della variabile attualmente selezionata che verrà utilizzata come base della nuova variabile categorizzata.

Variabile categorizzata. Nome ed etichetta facoltativa della nuova variabile categorizzata.

- **Nome.** È necessario specificare un nome per la nuova variabile. I nomi delle variabili devono essere univoci e rispettare le regole di denominazione delle variabili. Per ulteriori informazioni, vedere Nomi delle variabili in Capitolo 5 a pag. 79.
- **Etichetta.** È possibile specificare un'etichetta di variabile descrittiva contenente un massimo di 255 caratteri. L'etichetta di variabile predefinita corrisponde all'eventuale etichetta di variabile o nome di variabile della variabile sorgente con l'aggiunta di *(Categorizzato)* alla fine dell'etichetta.

Minimo e Massimo. I valori minimo e massimo della variabile attualmente selezionata, in base ai casi esaminati ed esclusi i valori definiti come valori mancanti definiti dall'utente.

Valori non mancanti. L'istogramma mostra la distribuzione dei valori non mancanti della variabile attualmente selezionata, in base ai casi esaminati.

- Dopo aver definito gli intervalli per la nuova variabile, nell'istogramma vengono visualizzate linee verticali per indicare i punti di divisione che delimitano gli intervalli.
- È possibile trascinare le linee dei punti di divisione in posizioni diverse sull'istogramma, modificando gli intervalli delle categorizzazioni.
- È possibile rimuovere gli intervalli trascinando le linee dei punti di divisione all'esterno dell'istogramma.

Nota: l'istogramma (in cui sono visualizzati valori non mancanti) e i valori minimo e massimo sono basati sui valori esaminati. Se nell'analisi non vengono inclusi tutti i casi, è possibile che la distribuzione effettiva non venga rappresentata in modo preciso, soprattutto se il file di dati è stato ordinato in base alla variabile selezionata. Se non viene esaminato alcun caso, non sarà disponibile alcuna informazione sulla distribuzione dei valori.

Griglia. Visualizza i valori che definiscono i punti finali superiori e le etichette di valore opzionali di ogni intervallo.

- **Valore.** Il valore che definisce i punti finali superiori di ogni intervallo. È possibile specificare valori o utilizzare l'opzione Crea punti divisione per creare automaticamente intervalli basati sui criteri selezionati. Per impostazione predefinita, viene automaticamente incluso un punto di divisione con il valore *MAX*. L'intervallo contiene tutti i valori non mancanti superiori agli altri punti di divisione. L'intervallo delimitato dal punto di divisione inferiore include tutti i valori non mancanti inferiori o uguali a tale valore (o solo inferiori al valore, a seconda delle modalità di definizione dei punti finali superiori).
- **Etichetta.** Opzionale, etichette descrittive per i valori della nuova variabile categorizzata. Dal momento che i valori della nuova variabile sono semplicemente interi sequenziali da 1 a *n*, le etichette descrittive che indicano ciò che il valore rappresenta possono rivelarsi molto utili. È possibile specificare le etichette o utilizzare l'opzione Crea etichette per crearle automaticamente.

Per eliminare un intervallo dalla griglia

- ▶ Fare clic con il pulsante destro del mouse sulla cella *Valore* o *Etichetta* relativa all'intervallo.
- ▶ Dal menu di scelta rapida, scegliere Elimina la riga.

Nota: se si elimina l'intervallo *MAX*, a tutti i casi con valori superiori all'ultimo valore di punto di divisione specificato verrà assegnato il valore mancante di sistema della nuova variabile.

Per eliminare tutte le etichette o tutti gli intervalli definiti

- ▶ Fare clic con il pulsante destro del mouse in un punto qualsiasi della griglia.
- ▶ Dal menu di scelta rapida, scegliere Elimina tutte le etichette o Elimina tutti i punti di divisione.

Punti finali superiori. Controlla il trattamento dei valori finali superiori nella colonna *Valore* della griglia.

- **Inclusi (<=).** I casi con il valore specificato nella cella *Valore* sono inclusi nella categoria categorizzata. Ad esempio, se si specificano come valori 25, 50 e 75, i casi con il valore 25 verranno inclusi nel primo intervallo, in quanto tale intervallo conterrà tutti i casi con valori inferiori o uguali a 25.
- **Esclusi (<).** I casi con il valore specificato nella cella *Valore* non sono inclusi nella categoria categorizzata, ma sono inclusi nell'intervallo successivo. Ad esempio, se si specificano i valori 25, 50 e 75, i casi con il valore 25 verranno inclusi nel secondo intervallo anziché nel primo, in quanto il primo intervallo conterrà tutti i casi con valori inferiori a 25.

Crea punti divisione. Genera automaticamente categorie raggruppate per intervalli di larghezza uguale, intervalli con lo stesso numero di casi o intervalli basati su deviazioni standard. Non è disponibile se non viene esaminato alcun caso. Per ulteriori informazioni, vedere Generazione automatica di categorie raggruppate a pag. 131.

Crea etichette. Genera etichette descrittive per i valori interi sequenziali della nuova variabile categorizzata, in base ai valori nella griglia e al trattamento specifico dei punti finali superiori (inclusi o esclusi).

Scala inversa. Per impostazione predefinita, i valori della nuova variabile categorizzata sono interi sequenziali crescenti da 1 a *n*. L'inversione della scala modifica tali valori in interi sequenziali decrescenti da *n* a 1.

Copia intervalli. È possibile copiare le specifiche di categorizzazione di un'altra variabile nella variabile attualmente selezionata o dalla variabile selezionata in più variabili diverse. Per ulteriori informazioni, vedere Copia di categorie raggruppate a pag. 133.

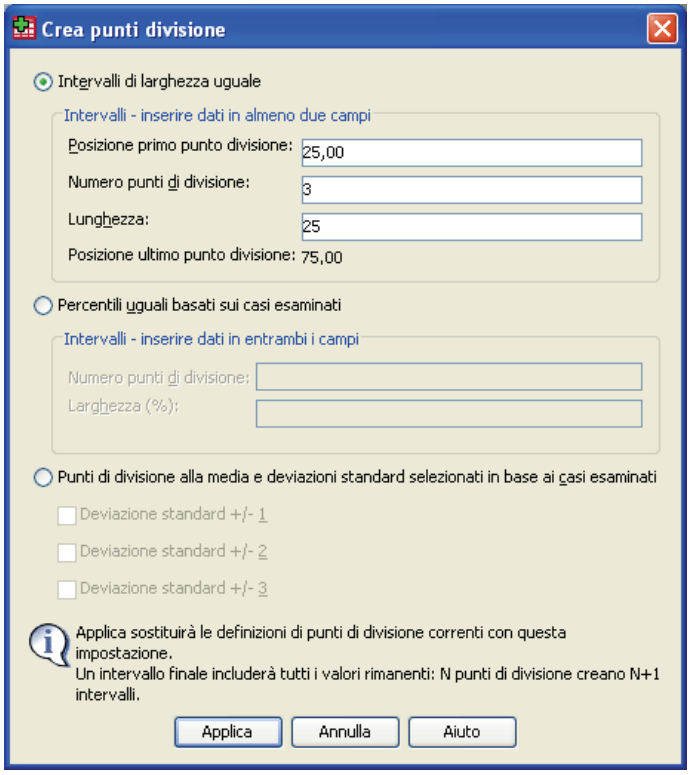
Generazione automatica di categorie raggruppate

La finestra di dialogo Crea punti divisione consente di generare automaticamente categorie raggruppate in base a criteri selezionati.

Per utilizzare la finestra di dialogo Crea punti divisione:

- Selezionare una variabile dall'Elenco variabili esaminate.
- Fare clic su Crea punti divisione.
- Selezionare i criteri per la generazione dei punti di divisione che definiranno le categorie raggruppate.
- Fare clic su Applica.

Figura 7-16
Finestra di dialogo Crea punti divisione


Crea punti divisione
Intervalli di larghezza uguale
Intervalli - inserire dati in almeno due campi
Posizione primo punto divisione:
25,00
Numero punti di divisione:
3
Lunghezza:
25
Posizione ultimo punto divisione:
75,00
Percentili uguali basati sui casi esaminati
Intervalli - inserire dati in entrambi i campi
Numero punti di divisione:
Larghezza (%):
Punti di divisione alla media e deviazioni standard selezionati in base ai casi esaminati
Deviazione standard +/- 1
Deviazione standard +/- 2
Deviazione standard +/- 3
Applica sostituirà le definizioni di punti di divisione correnti con questa impostazione.
Un intervallo finale includerà tutti i valori rimanenti: N punti di divisione creano N+1 intervalli.
Applica
Annulla
Aiuto

Nota: la finestra di dialogo Crea punti divisione non è disponibile se non viene esaminato alcun caso.

Intervalli di larghezza uguale. Genera categorie raggruppate di larghezza uguale (ad esempio, 1–10, 11–20 e 21–30) in base a due dei tre criteri seguenti:

- **Posiz. primo punto div.** Valore che definisce il punto superiore della categoria raggruppata inferiore (ad esempio, il valore 10 indica un intervallo che include tutti i valori fino a 10).
- **N. punti di divisione.** Il numero di categorie raggruppate corrisponde al numero di punti di divisione *più uno*. Ad esempio, 9 punti di divisione generano 10 categorie raggruppate.
- **Larghezza.** Larghezza di ogni intervallo. Ad esempio, il valore 10 determina la categorizzazione di *Età in anni* in intervalli di 10 anni.

Percentili uguali basati sui casi esaminati. Genera categorie raggruppate con un numero uguale di casi in ogni intervallo (mediante l'**algoritmo empirico** per i percentili), in base a uno dei criteri seguenti:

- **N. punti di divisione.** Il numero di categorie raggruppate corrisponde al numero di punti di divisione *più uno*. Ad esempio, tre punti di divisione generano quattro intervalli percentili (quartili) contenenti ognuno il 25% dei casi.
- **Larghezza (%).** Larghezza di ogni intervallo, espressa come percentuale del numero totale di casi. Ad esempio, il valore 33,3 determina la creazione di tre categorie raggruppate (due punti di divisione) contenenti ognuna il 33,3% dei casi.

Se la variabile sorgente contiene un numero relativamente limitato di valori distinti o un numero di casi con lo stesso valore, è possibile che venga creato un numero di intervalli inferiore a quello richiesto. Se esistono più valori identici per un punto di divisione, questi vengono inseriti tutti nello stesso intervallo; quindi è possibile che le percentuali effettive non siano sempre esattamente uguali.

Punti di divisione alla media e deviaz. stand. selezionati in base ai casi esaminati. Genera categorie raggruppate basate sui valori di media e deviazione standard della distribuzione della variabile.

- Se non si seleziona nessuno degli intervalli di deviazione standard, verranno create due categorie raggruppate che utilizzano la media come punto di divisione.
- È possibile selezionare qualsiasi combinazione di intervalli di deviazione standard basata su una, due e/o tre deviazioni standard. Ad esempio, la selezione di tutte e tre determina la creazione di otto categorie raggruppate, ovvero sei categorizzazioni negli intervalli di una deviazione standard e due categorizzazioni per i casi superiori alle tre deviazioni sopra e sotto la media.

In una distribuzione normale, il 68% dei casi è compreso in un intervallo pari a una deviazione standard della media, il 95% in un intervallo pari a due deviazioni standard e il 99% in un intervallo pari a tre deviazioni standard. La creazione di categorie raggruppate basate su deviazioni standard può determinare la creazione di intervalli definiti esterni all'intervallo di dati effettivo o anche esterni all'intervallo di valori di dati possibili (ad esempio, in un intervallo di stipendi negativo).

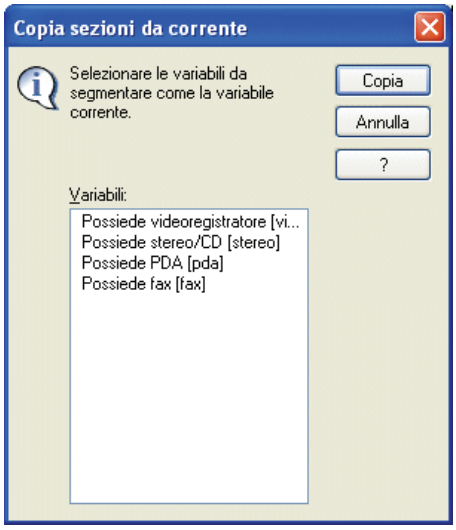
Nota: i calcoli dei percentili e delle deviazioni standard sono basati sui casi esaminati. Se si limita il numero di casi esaminati, è possibile che gli intervalli risultanti non contengano la proporzione di casi desiderata in tali intervalli, soprattutto se il file di dati è ordinato in base alla variabile

sorgente. Ad esempio, se si limita l'analisi ai primi 100 casi di un file di dati con 1000 casi e il file di dati viene ordinato in base all'ordine crescente di età dei rispondenti, anziché categorizzazioni percentili di età contenenti ognuna il 25% dei casi è possibile che i primi tre intervalli creati contengano ognuno circa il 3,3% dei casi e l'ultimo contenga il 90% dei casi.

Copia di categorie raggruppate

Quando si creano categorie raggruppate per una o più variabili, è possibile copiare le specifiche di categorizzazione di un'altra variabile nella variabile attualmente selezionata o dalla variabile selezionata in più variabili diverse.

Figura 7-17
Copia di intervalli nella variabile corrente o dalla variabile corrente



Per copiare specifiche di categorizzazione

- Definire categorie raggruppate per almeno una variabile ma *non* fare clic su OK o su Incolla.
- Selezionare una variabile dall'Elenco variabili esaminate per cui sono state definite categorie raggruppate.
- Fare clic su Ad altre variabili.
- Selezionare le variabili per cui si desidera creare nuove variabili con le stesse categorie raggruppate.
- Fare clic su Copia.

o

- Selezionare una variabile dall'Elenco variabili esaminate in cui si desidera copiare categorie raggruppate definite.
- Fare clic su Da un'altra variabile.
- Selezionare la variabile con le categorie raggruppate definite che si desidera copiare.

▶ Fare clic su Copia.

Se per la variabile da cui si copiano le specifiche di categorizzazione sono state specificate etichette di valori, verranno copiate anche tali etichette.

Nota: dopo aver fatto clic su OK nella finestra principale Categorizzazione visuale per creare nuove variabili categorizzate (o dopo aver chiuso tale finestra in altro modo), non sarà possibile utilizzare Categorizzazione visuale per copiare le categorie raggruppate in altre variabili.

Valori mancanti definiti dall'utente in Categorizzazione visuale

I valori definiti come valori mancanti definiti dall'utente (ovvero i valori identificati come codici per dati mancanti) per la variabile sorgente non sono inclusi nelle categorie raggruppate per la nuova variabile. I valori mancanti definiti dall'utente per le variabili sorgente vengono copiati come valori mancanti definiti dall'utente della nuova variabile insieme a tutte le etichette di valori definiti per i codici di valori mancanti.

Se un codice di valore mancante è in conflitto con uno dei valori delle categorie raggruppate per la nuova variabile, il conflitto viene risolto utilizzando per la registrazione di tale codice nella nuova variabile un valore corrispondente all'aggiunta di 100 al valore più alto delle categorie raggruppate. Ad esempio, se il valore 1 è definito come valore mancante definito dall'utente per la variabile sorgente e la nuova variabile contiene sei categorie raggruppate, tutti i casi con valore 1 avranno valore 106 per la nuova variabile e 106 verrà definito come valore mancante definito dall'utente. Se il valore mancante definito dall'utente per la variabile sorgente include un'etichetta di valore definita, tale etichetta verrà mantenuta come etichetta del valore registrato per la nuova variabile.

Nota: se la variabile sorgente include un intervallo definito di valori mancanti definiti dall'utente nel formato *LO-n*, dove *n* corrisponde a un numero positivo, i valori mancanti definiti dall'utente corrispondenti per la nuova variabile saranno numeri negativi.

Capitolo
8

Trasformazioni di dati

Nella situazione ideale, i dati sono perfettamente adatti al tipo di analisi che si desidera eseguire e le relazioni tra le variabili sono lineari oppure nettamente ortogonali. Purtroppo questo caso non si verifica quasi mai. Le analisi preliminari possono rivelare la presenza di schemi di codifica difficili o di errori di codifica, oppure può essere necessario eseguire trasformazioni di dati per identificare la relazione effettiva tra le variabili.

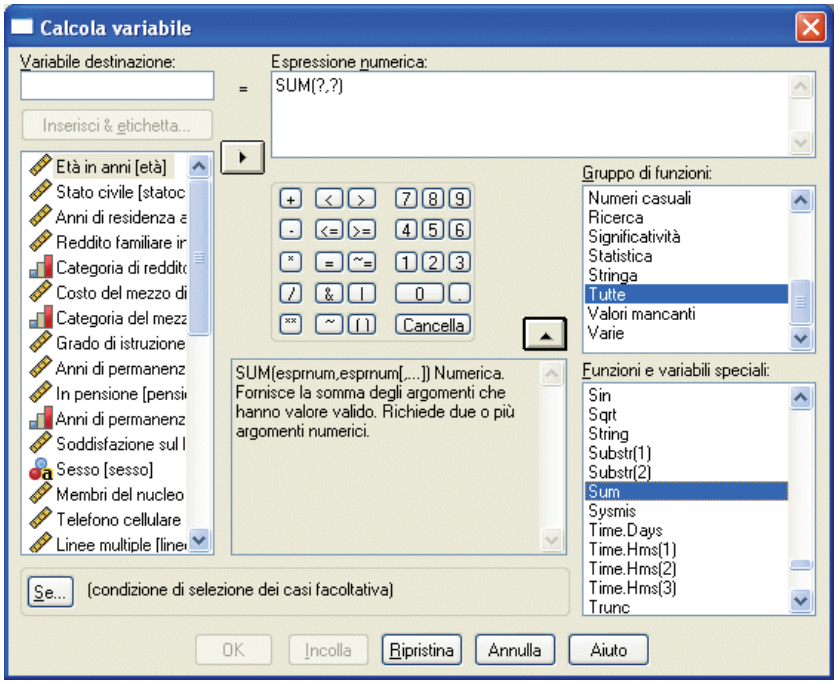
È possibile eseguire trasformazioni di dati rappresentate da attività semplici, ad esempio la compressione delle categorie per l'analisi, o da attività più complesse, come la creazione di nuove variabili in base ad equazioni complesse e istruzioni condizionali.

Calcolo delle variabili

Utilizzare la finestra di dialogo Calcola per calcolare i valori di una variabile basata su trasformazioni numeriche di altre variabili.

- È possibile calcolare i valori di variabili numeriche o variabili stringa (alfanumeriche).
- È possibile creare nuove variabili oppure sostituire i valori di variabili esistenti. Per le nuove variabili è inoltre possibile specificare il tipo e l'etichetta.
- È possibile calcolare i valori selettivamente per sottoinsiemi di dati in base a condizioni logiche.
- È possibile utilizzare più di 70 funzioni predefinite, ad esempio funzioni aritmetiche, statistiche, di distribuzione e stringa.

Figura 8-1
Finestra di dialogo Calcola variabile



Per calcolare variabili

- Dai menu, scegliere:

  Trasforma
  Calcola variabile...

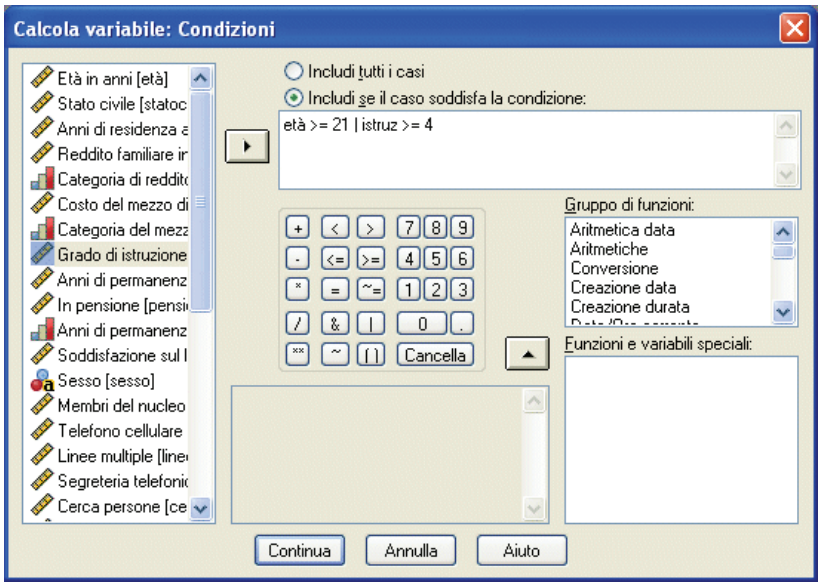
- Digitare il nome di una sola variabile di destinazione. È possibile aggiungere all'insieme di dati attivo una variabile nuova o esistente.

- Per creare un'espressione, è possibile incollare o digitare direttamente i componenti nel campo Espressione.
  - È possibile incollare le funzioni o le variabili di sistema normalmente utilizzate selezionando un gruppo nell'elenco delle funzioni e facendo doppio clic su una funzione o variabile nell'elenco delle funzioni e variabili speciali. È possibile anche selezionare la funzione o la variabile, quindi fare clic sulla freccia vicino all'elenco delle funzioni. Immettere i valori per tutti i parametri contrassegnati con punti interrogativi (solo per le funzioni). Il gruppo di funzioni denominato Tutti contiene un elenco di tutte le funzioni e variabili di sistema disponibili. L'area dedicata alla finestra dialogo visualizza una breve descrizione della funzione o variabile correntemente selezionata.
  - Le costanti stringa devono essere incluse tra virgolette o apostrofi.
  - Se i valori contengono decimali, usare il punto (.) come separatore decimale.
  - Per le nuove variabili stringa, è necessario selezionare anche Tipo ed etichetta per specificare il tipo di dati.

Calcola variabile: Condizioni

Nella finestra di dialogo Condizioni è possibile applicare trasformazioni di dati al sottoinsieme di casi selezionato utilizzando le espressioni logiche. Un'espressione logica restituisce un valore *vero*, *falso*, o *mancante* per ciascun caso.

Figura 8-2
Finestra di dialogo Calcola variabile: Condizioni



- Se il risultato di un'espressione logica è *vero*, il caso verrà incluso nel sottoinsieme selezionato.
- Se il risultato di un'espressione logica è *falso* o *mancante*, il caso non viene incluso nel sottoinsieme selezionato.
- La maggior parte delle espressioni condizionali utilizza uno o più dei sei operatori relazionali (<, >, <=, >=, = e ~=) della calcolatrice.
- Le espressioni logiche possono includere nomi di variabili, costanti, operatori aritmetici, funzioni numeriche e di altro tipo, variabili logiche e operatori relazionali.

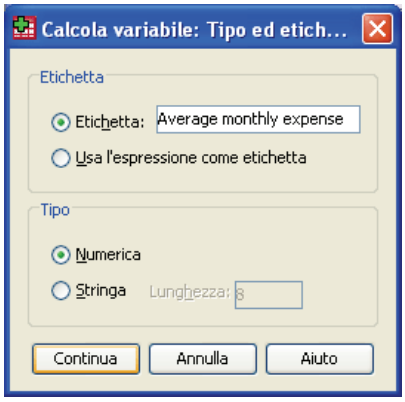
Calcola variabile: Tipo ed etichetta

Per impostazione predefinita, le nuove variabili calcolate sono numeriche. Per calcolare una nuova variabile stringa, è necessario specificare il tipo di dati e la larghezza.

Etichetta. Etichetta di variabile descrittiva e opzionale, con una lunghezza massima di 255 byte. È possibile immettere un'etichetta oppure utilizzare i primi 110 caratteri dell'espressione Calcola come etichetta.

Tipo. Le variabili calcolate possono essere di tipo numerico o stringa (alfanumeriche). Nei calcoli non è possibile utilizzare le variabili stringa.

Figura 8-3
Finestra di dialogo Tipo ed etichetta.



Funzioni

Sono supportati molti tipi di funzioni, inclusi:

- Funzioni aritmetiche
- Funzioni statistiche
- Funzioni stringa
- Funzioni di data e di ora
- Funzioni di distribuzione
- Funzioni per variabili casuali
- Funzioni per valori mancanti
- Funzioni di punteggio (solo per SPSS Statistics Server)

Per ulteriori informazioni e per una dettagliata descrizione di ciascuna funzione, digitare funzioni nella scheda Indice della Guida.

■

Valori mancanti nelle funzioni

Nelle funzioni e nelle espressioni matematiche semplici i valori mancanti vengono trattati in modo diverso. Nell'espressione seguente:

```
(var1+var2+var3)/3
```

il risultato è mancante se un caso include un valore mancante per una delle tre variabili.

Nell'espressione seguente:

```
MEAN(var1, var2, var3)
```

il risultato è mancante solo se il caso include valori mancanti per tutte e tre le variabili.

Per quanto riguarda le funzioni statistiche, è possibile specificare il numero minimo di argomenti che devono includere valori non mancanti. Per eseguire questa operazione, digitare un punto e il numero minimo dopo il nome della funzione, ad esempio:

```
MEAN.2(var1, var2, var3)
```

Generatori di numeri casuali

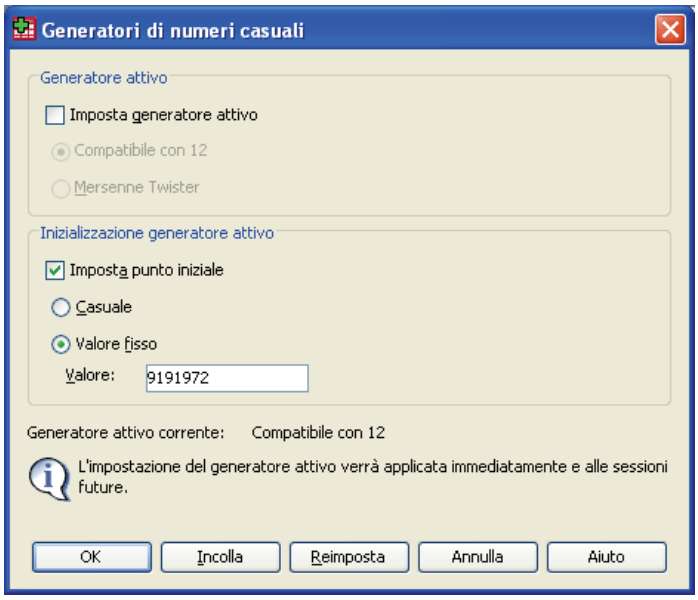
La finestra di dialogo Generatore di numeri casuale permette di selezionare il generatore di numeri casuale e di impostare il valore iniziale della sequenza da utilizzare per riprodurre una sequenza di numeri casuali.

Generatore attivo. Sono disponibili due diversi generatori di numeri casuali:

- **Compatibile con la versione 12.** Il generatore di numeri casuali utilizzato nella versione 12 e precedenti. Utilizzare questo generatore se è necessario riprodurre i risultati randomizzati generati in versioni precedenti in base a un valore di seme specificato.
- **Mersenne Twister.** Un generatore di numeri casuali più affidabile a scopi di simulazione. Se la riproduzione di risultati randomizzati dalla versione 12 o precedenti non rappresenta un problema, utilizzare questo generatore.

Inizializzazione generatore attivo. Il seme dei numeri casuali cambia ogni volta che viene generato un numero casuale da utilizzare per le trasformazioni (ad esempio le funzioni di distribuzione casuale), il campionamento casuale e la ponderazione dei casi. Per replicare una sequenza di numeri casuali, impostare il valore del punto di inizializzazione iniziale prima di eseguire l'analisi basata su numeri casuali. Il valore deve essere un numero positivo intero.

Figura 8-4
Finestra di dialogo Generatori di numeri casuali



Per selezionare il generatore di numeri casuale e/o impostare il valore di inizializzazione:

▶ Dai menu, scegliere:

Trasforma
Generatori di numeri casuali

Conteggia occorrenze di valori all'interno dei casi

Questa finestra di dialogo consente di creare una variabile per il conteggio delle occorrenze dello stesso valore in una lista di variabili relative allo stesso caso. Ad esempio, un'indagine può contenere una lista di riviste con le caselle *sì*/*no* in cui è possibile indicare quali riviste vengono lette dai rispondenti. È possibile specificare il numero di risposte *affermative* fornite da ciascuna persona per creare una nuova variabile che contenga il numero totale di riviste lette.

Figura 8-5
Finestra di dialogo Conteggia occorrenze di valori all'interno dei casi

Per contare le occorrenze dei valori all'interno dei casi

▶ Dai menu, scegliere:

Trasforma
Conta valori all'interno dei casi...

▶ Specificare il nome della variabile di destinazione.

▶ Selezionare due o più variabili dello stesso tipo (numerico o stringa).

▶ Fare clic su Definisci valori e specificare i valori da contare.

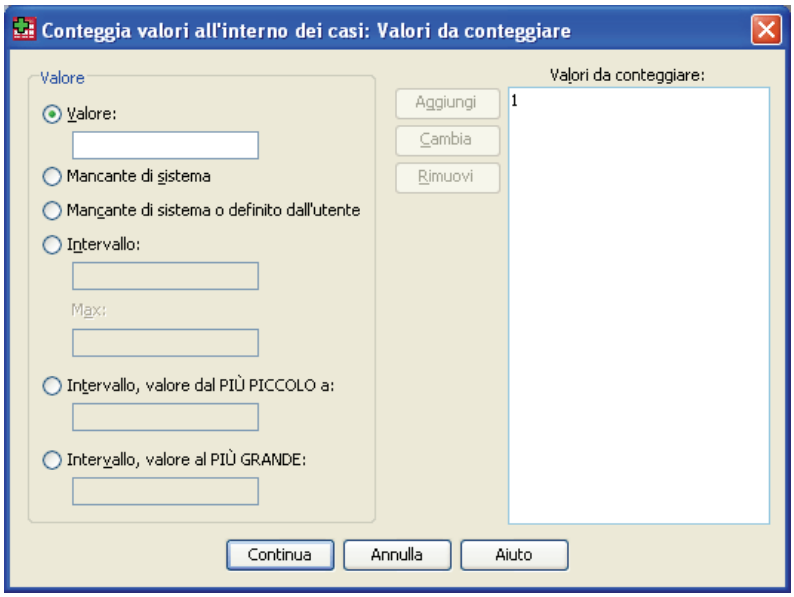
È inoltre possibile definire un sottoinsieme di casi in cui contare le occorrenze dei valori.

Conteggia valori all'interno dei casi: Valori da conteggiare

Il valore della variabile di destinazione (nella finestra di dialogo principale) aumenta di 1 ogni volta che le variabili selezionate corrispondono a una specificazione della lista Valori da conteggiare. Se un caso corrisponde a più specificazioni per la stessa variabile, la variabile di destinazione aumenterà più volte per tale variabile.

Le specificazioni dei valori possono includere valori singoli, valori mancanti o valori mancanti di sistema e intervalli. Gli intervalli includono i relativi punti finali e i valori mancanti definiti dall'utente compresi in ciascun intervallo.

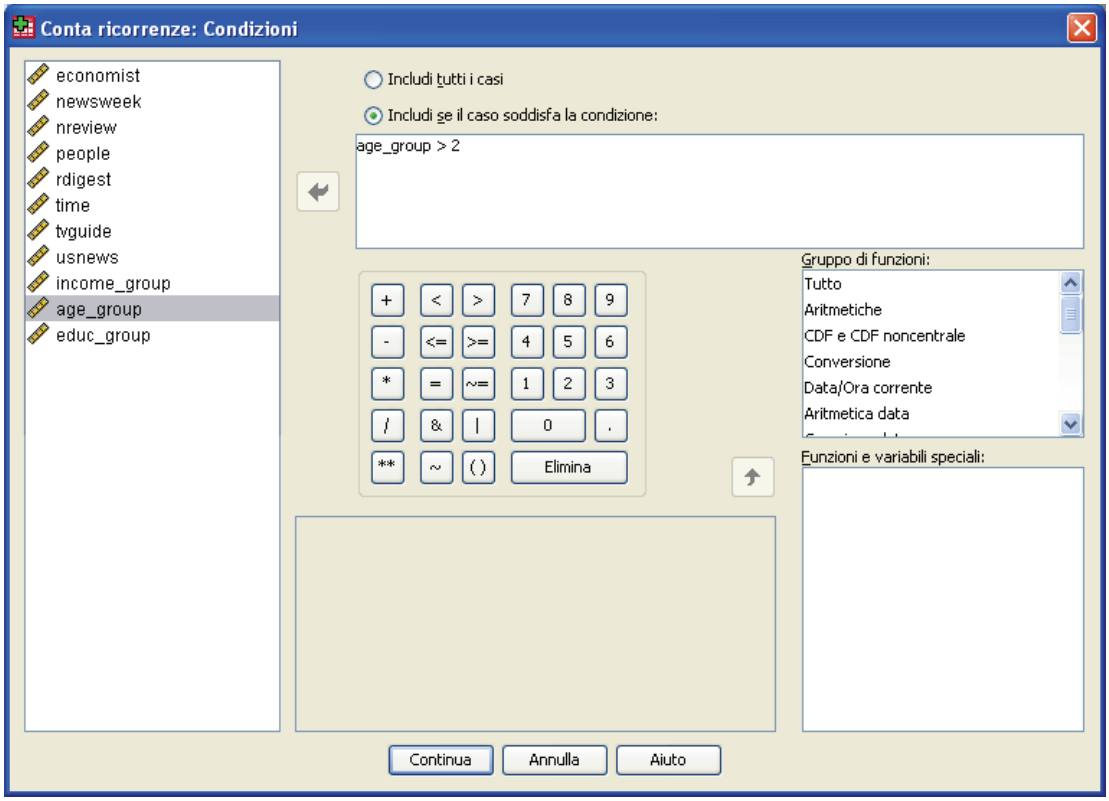
Figura 8-6
Finestra di dialogo Valori da conteggiare



Conta ricorrenze: Condizioni

La finestra di dialogo Condizioni permette di contare le ricorrenze dei valori di un sottoinsieme specifico di casi mediante espressioni condizionali. Un'espressione logica restituisce un valore *vero*, *falso*, o *mancante* per ciascun caso.

Figura 8-7
Finestra di dialogo Conta ricorrenze: Condizioni



Per informazioni generali sull'uso della finestra di dialogo Condizioni, vedere Calcola variabile: Condizioni a pag. 137.

Sposta valori

Sposta valori crea nuove variabili contenenti i valori di variabili esistenti in casi precedenti o successivi.

Nome. Nome della nuova variabile. Deve essere un nome che non esiste già nell'insieme di dati attivo.

Ottieni valore da un caso precedente (ritardo). Ottiene il valore da un caso precedente nell'insieme di dati attivo. Ad esempio, con il valore del numero predefinito di casi 1, ogni caso per la nuova variabile ottiene il valore della variabile originale dal caso immediatamente precedente.

Ottieni valore da caso successivo (anticipo). Ottiene il valore da un caso successivo nell'insieme di dati attivo. Ad esempio, con valore del numero predefinito di casi 1, ogni caso per la nuova variabile ottiene il valore della variabile originale dal caso successivo.

Numero di casi da spostare. Ottiene il valore dal *n*° caso precedente o successivo, dove *n* è il valore specificato. Il valore deve essere un numero intero non negativo.

- Se è in corso l'analisi del file, l'ambito dello spostamento si limita a ogni gruppo di analisi. Non è possibile ottenere un valore di spostamento da un caso che si trova in un gruppo di analisi precedente o successivo.
- Lo stato del filtro viene ignorato.
- Il valore della variabile risultante viene impostato su mancante di sistema per i primi o gli ultimi *n* casi nell'insieme di dati o blocco di dati, dove *n* è il valore specificato per **Numero di casi da spostare**. Ad esempio, se si utilizza il metodo con ritardo con valore 1, la variabile risultante verrà impostata su mancante di sistema per il primo caso dell'insieme di dati (o il primo caso in ciascun blocco di dati).
- I valori mancanti definiti dall'utente vengono conservati.
- Le informazioni del dizionario relative alla variabile originale, comprese le etichette di valori definite e le assegnazioni dei valori mancanti definiti dall'utente, vengono applicate alla nuova variabile. (Nota: gli attributi delle variabili personalizzati non vengono inclusi).
- Viene automaticamente generata un'etichetta di variabile per la nuova variabile che descrive l'operazione di spostamento che ha creato la variabile.

Per creare una nuova variabile con i valori spostati

▶ Dai menu, scegliere:
Trasforma
Sposta valori

▶ Selezionare la variabile da utilizzare come origine dei valori per la nuova variabile.

▶ Specificare un nome per la nuova variabile.

▶ Selezionare il metodo di spostamento (ritardo o anticipo) e il numero di casi da spostare.

▶ Fare clic su Cambia.

▶ Ripetere l'operazione per ciascuna nuova variabile da creare.

Ricodifica dei valori

Per modificare i valori, è possibile ricodificarli. Questo risulta particolarmente utile per comprimere o combinare più categorie. È possibile ricodificare i valori all'interno delle variabili esistenti oppure creare nuove variabili in base ai valori ricodificati delle variabili esistenti.

Ricodifica nelle stesse variabili

La finestra di dialogo Ricodifica nelle stesse variabili permette di riassegnare i valori delle variabili esistenti o di comprimere gli intervalli dei valori esistenti in nuovi valori. È possibile, ad esempio, comprimere i salari in categorie di intervalli di salari.

È possibile ricodificare variabili numeriche e variabili stringa. Se si selezionano più variabili, queste devono essere dello stesso tipo. Non è possibile ricodificare contemporaneamente variabili stringa e numeriche.

Figura 8-8
Finestra di dialogo Ricodifica nelle stesse variabili

Per ricodificare i valori di una variabile

- Dai menu, scegliere:
  Trasforma
  Ricodifica nelle stesse variabili...

- Selezionare le variabili che si desidera ricodificare. Se vengono selezionate più variabili, devono essere dello stesso tipo (numerico o stringa).

- Fare clic su Valori vecchi e nuovi e specificare le modalità di ricodifica dei valori.

  È inoltre possibile definire un sottoinsieme di casi da ricodificare. La finestra di dialogo Condizioni è la stessa di quella descritta per Conta occorrenze.

Ricodifica nelle stesse variabili: Valori vecchi e nuovi

In questa finestra di dialogo è possibile definire i valori da ricodificare. Tutte le specificazioni dei valori devono essere dello stesso tipo di dati (numerico o stringa) delle variabili selezionate nella finestra di dialogo principale.

Vecchio valore. I valori da ricodificare. È possibile ricodificare singoli valori, intervalli di valori o valori mancanti. I valori e gli intervalli mancanti di sistema non possono essere selezionati come variabili stringa in quanto nessuno dei due concetti è valido per tali variabili. Gli intervalli includono i relativi punti finali e i valori mancanti definiti dall'utente compresi in ciascun intervallo.

- **Valore.** Valore discreto da ricodificare in un nuovo valore. Il valore deve essere dello stesso tipo (numerico o stringa) delle variabili da ricodificare.
- **Mancante di sistema.** Il valore mancante di sistema rappresenta dati numerici non noti. Viene usato come risultato di calcoli che non possono essere eseguiti ed è assegnato quando campi completamente vuoti vengono letti come dati numerici. Il valore mancante di sistema viene rappresentato tramite un punto. Questo tipo di valore non esiste nelle variabili stringa in quanto qualsiasi carattere è valido in queste variabili.

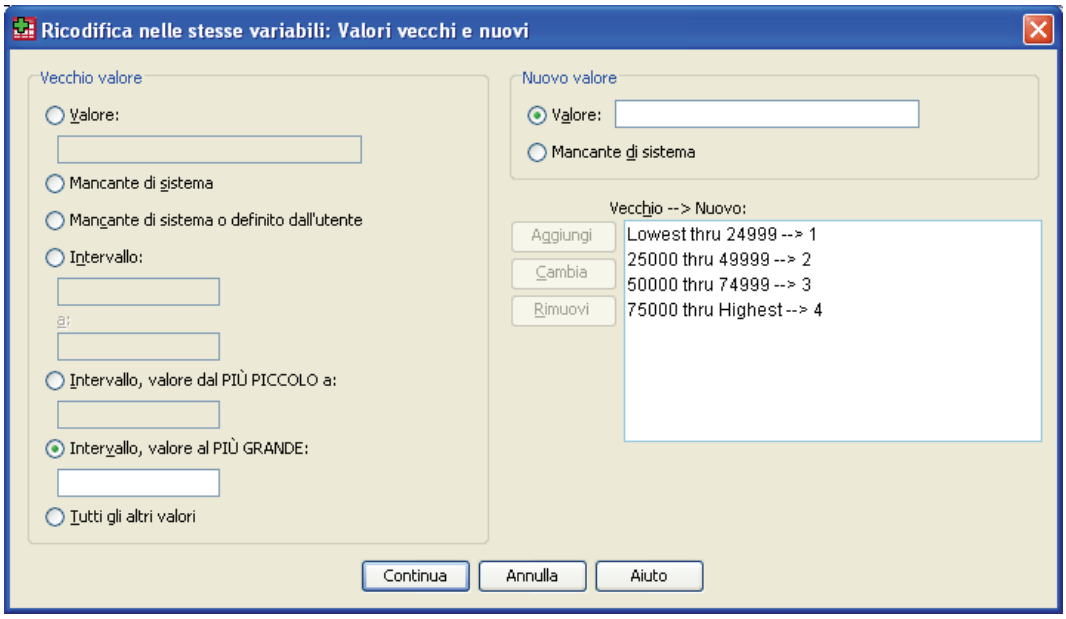
- **Mancante di sistema o definito dall'utente.** Osservazioni con valori che sono stati dichiarati mancanti dall'utente oppure non noti (mancanti di sistema), indicati da un punto (.).
- **Intervallo.** Un intervallo di valori, inclusi gli estremi specificati. Non disponibile per variabili stringa. Sono inclusi tutti i valori mancanti definibili dall'utente compresi nell'intervallo.
- **Tutti gli altri valori.** Tutti i valori restanti non inseriti in una delle specifiche nell'elenco delle ricodifiche. Aggiunge una clausola ELSE all'elenco delle ricodifiche.

Nuovo valore. Il singolo valore in cui viene ricodificato ciascun vecchio valore o intervallo di valori. È possibile specificare un valore oppure assegnare il valore mancante di sistema.

- **Valore.** Valore destinato a sostituire uno o più vecchi valori. Deve essere dello stesso tipo (numerico o stringa) del vecchio valore corrispondente.
- **Mancante di sistema.** Sostituisce al vecchio valore specificato il valore mancante di sistema. I valori mancanti vengono esclusi dalla maggior parte delle analisi. Non disponibile per variabili stringa.

Vecchio–>Nuovo. La lista delle specificazioni che verranno utilizzate per ricodificare le variabili. È possibile aggiungere, modificare e rimuovere le specifiche dall'elenco. L'elenco verrà ordinato automaticamente in base alla specifiche del vecchio valore e con il seguente ordine: valori unici, valori mancanti, intervalli e tutti gli altri valori. Se nell'elenco viene modificata una specifica di ricodifica, la procedura riordina automaticamente l'elenco, se necessario, per mantenere tale ordine.

Figura 8-9
Finestra di dialogo Valori vecchi e nuovi

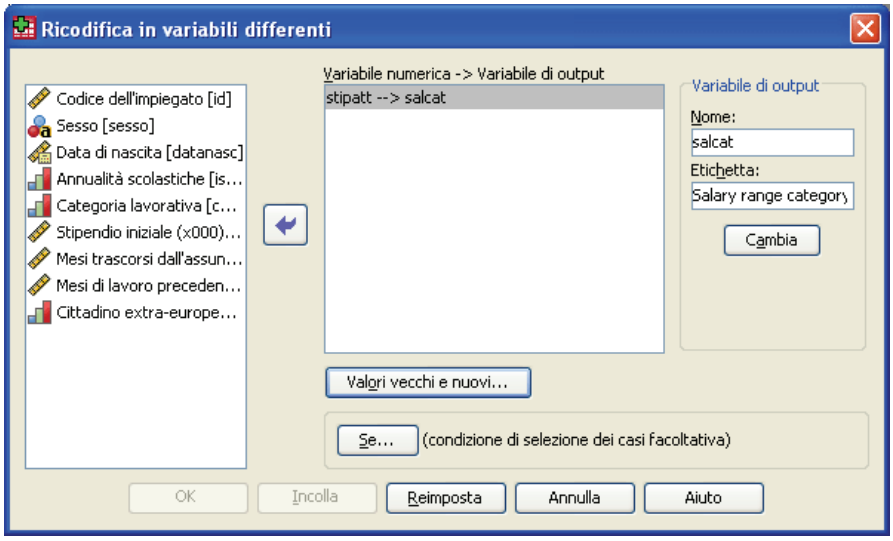


Ricodifica in variabili differenti

La finestra di dialogo Ricodifica in variabili differenti permette di riassegnare i valori delle variabili esistenti o di comprimere gli intervalli dei valori esistenti nei nuovi valori della variabile. È possibile, ad esempio, comprimere i salari in una nuova variabile che contenga categorie di intervalli di salari.

- È possibile ricodificare variabili numeriche e variabili stringa.
- È possibile ricodificare variabili numeriche in variabili stringa e viceversa.
- Se si selezionano più variabili, queste devono essere dello stesso tipo. Non è possibile ricodificare contemporaneamente variabili stringa e numeriche.

Figura 8-10
Finestra di dialogo Ricodifica in variabili differenti



Per ricodificare i valori di una variabile in una nuova variabile

▶ Dai menu, scegliere:
Trasforma
Ricodifica in variabili differenti...

▶ Selezionare le variabili che si desidera ricodificare. Se vengono selezionate più variabili, devono essere dello stesso tipo (numerico o stringa).

▶ Specificare un (nuovo) nome di variabile di output per ciascuna nuova variabile e fare clic su Cambia.

▶ Fare clic su Valori vecchi e nuovi e specificare le modalità di ricodifica dei valori.

È inoltre possibile definire un sottoinsieme di casi da ricodificare. La finestra di dialogo Condizioni è la stessa di quella descritta per Conta occorrenze.

Ricodifica in variabili differenti: Valori vecchi e nuovi

In questa finestra di dialogo è possibile definire i valori da ricodificare.

Vecchio valore. I valori da ricodificare. È possibile ricodificare singoli valori, intervalli di valori o valori mancanti. I valori e gli intervalli mancanti di sistema non possono essere selezionati come variabili stringa in quanto nessuno dei due concetti è valido per tali variabili. I vecchi valori devono avere lo stesso tipo di dati (numerico o stringa) della variabile originale. Gli intervalli includono i relativi punti finali e i valori mancanti definiti dall'utente compresi in ciascun intervallo.

- **Valore.** Valore discreto da ricodificare in un nuovo valore. Il valore deve essere dello stesso tipo (numerico o stringa) delle variabili da ricodificare.
- **Mancante di sistema.** Il valore mancante di sistema rappresenta dati numerici non noti. Viene usato come risultato di calcoli che non possono essere eseguiti ed è assegnato quando campi completamente vuoti vengono letti come dati numerici. Il valore mancante di sistema viene rappresentato tramite un punto. Questo tipo di valore non esiste nelle variabili stringa in quanto qualsiasi carattere è valido in queste variabili.
- **Mancante di sistema o definito dall'utente.** Osservazioni con valori che sono stati dichiarati mancanti dall'utente oppure non noti (mancanti di sistema), indicati da un punto (.).
- **Intervallo.** Un intervallo di valori, inclusi gli estremi specificati. Non disponibile per variabili stringa. Sono inclusi tutti i valori mancanti definibili dall'utente compresi nell'intervallo.
- **Tutti gli altri valori.** Tutti i valori restanti non inseriti in una delle specifiche nell'elenco delle ricodifiche. Aggiunge una clausola ELSE all'elenco delle ricodifiche.

Nuovo valore. Il singolo valore in cui viene ricodificato ciascun vecchio valore o intervallo di valori. I nuovi valori possono essere di tipo numerico o stringa.

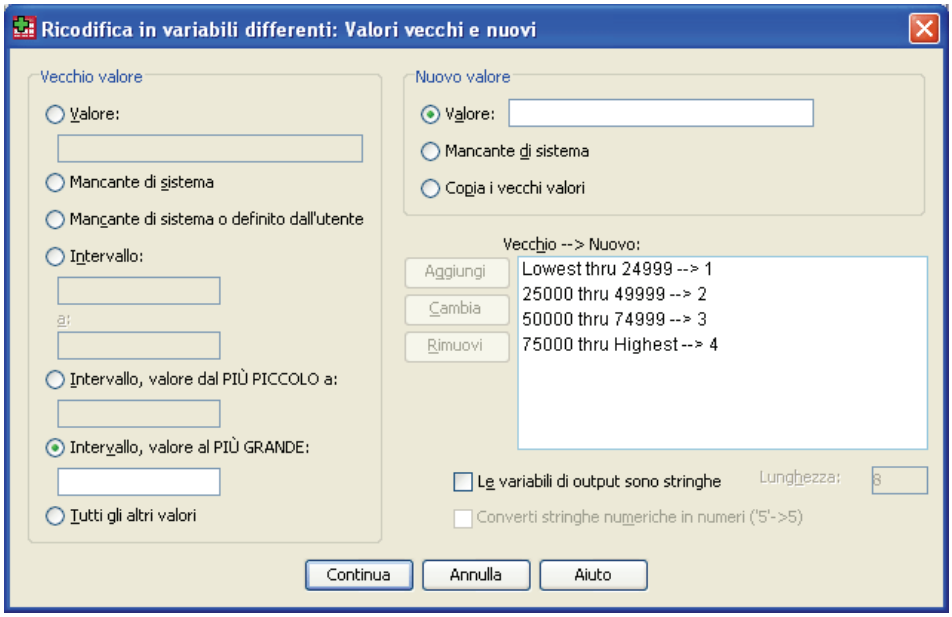
- **Valore.** Valore destinato a sostituire uno o più vecchi valori. Deve essere dello stesso tipo (numerico o stringa) del vecchio valore corrispondente.
- **Mancante di sistema.** Sostituisce al vecchio valore specificato il valore mancante di sistema. I valori mancanti vengono esclusi dalla maggior parte delle analisi. Non disponibile per variabili stringa.
- **Copia i vecchi valori.** Mantiene i vecchi valori, quando non ne è richiesta la ricodifica. Nella nuova variabile non verranno inclusi eventuali vecchi valori non specificati e ai casi che comprendono tali valori verrà assegnato il valore mancante di sistema per la nuova variabile.

Le variabili di output sono stringhe (Ricodifica). Definisce le nuove variabili ricodificate come stringhe. Le variabili originali possono essere numeriche o stringhe.

Converti stringhe numeriche in numeri. Converte in numeri le stringhe contenenti valori numerici. I casi con valori non numerici assumeranno valore mancante di sistema nella variabile di destinazione.

Vecchio–>Nuovo. La lista delle specificazioni che verranno utilizzate per ricodificare le variabili. È possibile aggiungere, modificare e rimuovere le specifiche dall'elenco. L'elenco verrà ordinato automaticamente in base alla specifiche del vecchio valore e con il seguente ordine: valori unici, valori mancanti, intervalli e tutti gli altri valori. Se nell'elenco viene modificata una specifica di ricodifica, la procedura riordina automaticamente l'elenco, se necessario, per mantenere tale ordine.

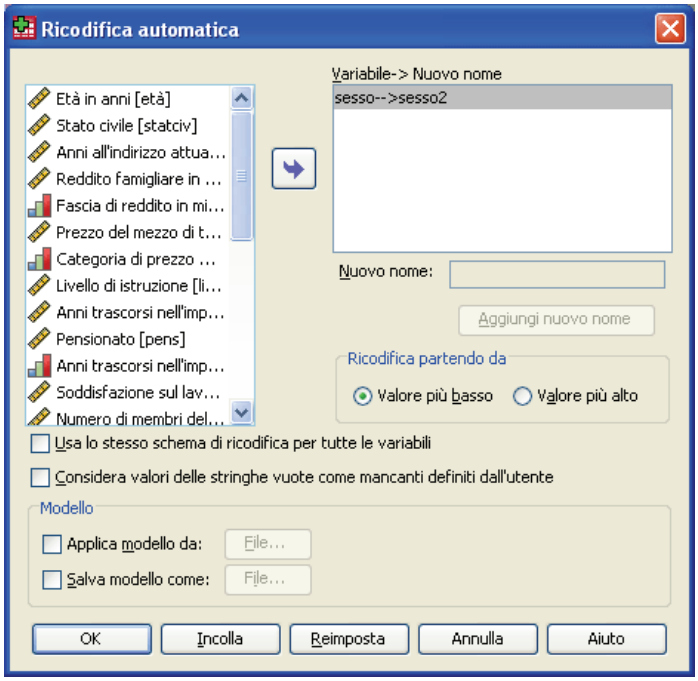
Figura 8-11
Finestra di dialogo Valori vecchi e nuovi



Ricodifica automatica

La finestra di dialogo Ricodifica automatica permette di convertire i valori stringa e numerici in numeri interi consecutivi. Se i codici di categoria non sono sequenziali, le celle vuote risultanti comportano prestazioni inferiori e maggiori requisiti di memoria per l'esecuzione di numerose procedure. Inoltre, in alcune procedure non è possibile utilizzare variabili stringa e altre richiedono valori interi consecutivi per i livelli del fattore.

Figura 8-12
Finestra di dialogo Ricodifica automatica



- Le nuove variabili create dalla procedura Ricodifica automatica mantengono le etichette delle variabili e dei valori definite per la vecchia variabile. Se per alcuni valori non è stata definita un'etichetta dei valori, come etichetta del valore ricodificato verrà utilizzato il valore originale. I valori vecchi e nuovi e le etichette dei valori verranno visualizzati in una tabella.
- I valori stringa verranno ricodificati in ordine alfabetico; le lettere maiuscole precedono le minuscole.
- I valori mancanti verranno ricodificati come maggiori dei valori non mancanti; il loro ordine verrà mantenuto. Se, ad esempio, la variabile originale ha 10 valori non mancanti, il valore mancante minimo verrà ricodificato a 11 e il valore 11 verrà utilizzato come valore mancante per la nuova variabile.

Usa lo stesso schema di ricodifica per tutte le variabili. Questa opzione permette di applicare lo stesso schema di ricodifica a tutte le variabili selezionate e di utilizzare quindi uno schema di codifica uniforme per tutte le nuove variabili.

Se si seleziona questa opzione, si applicano le seguenti regole e limitazioni:

- Tutte le variabili devono essere dello stesso tipo (numeriche o stringa).
- Tutti i valori osservati di tutte le variabili selezionate vengono utilizzate per creare un ordinamento dei valori utilizzabile per la ricodifica nei numeri interi sequenziali.
- I valori mancanti definiti dall'utente delle nuove variabili si basano sulla prima variabile dell'elenco dei valori mancanti definiti dall'utente. Tutti gli altri valori derivanti dalle variabili originali, ad eccezione di quelli mancanti definiti dal sistema, sono considerati validi.

Considera i valori delle stringhe vuote come mancanti definiti dall'utente. Per le variabili stringa, i valori vuoti o nulli non vengono considerati come valori **mancanti di sistema**. Questa opzione eseguirà la ricodifica automatica delle stringhe vuote in un valore **mancante definito dall'utente** superiore al valore più elevato non mancante.

Modelli

È possibile salvare lo schema di ricodifica in un file del modello e applicarlo a tutte le altre variabili o agli altri file di dati.

Esempio: si devono ricodificare molti codici prodotto alfanumerici in numeri interi ogni mese ed aggiungere, in alcuni mesi, nuovi codici prodotto che modificano lo schema di ricodifica automatica originale. Se si salva lo schema originale in un modello e lo si applica ai nuovi dati che contengono il nuovo insieme di codici, tutti i nuovi codici contenuti nei dati vengono automaticamente ricodificati in valori maggiori dell'ultimo valore del modello, senza che venga modificato lo schema di ricodifica automatica dei codici prodotto originali.

Salva modello con nome. Salva lo schema di ricodifica automatico per le variabili selezionate in un file del modello esterno.

- Questo modello contiene informazioni sulle associazioni tra i valori non mancanti originali ed i valori ricodificati.
- Nel modello vengono salvate solo le informazioni relative ai valori non mancanti. Le informazioni relative ai valori non mancanti specificati dall'utente non vengono salvate.
- Se sono state selezionate più variabili per la ricodifica ma non è stato selezionato lo stesso schema di ricodifica automatica per tutte le variabili o non è prevista l'applicazione di un modello esistente allo schema di ricodifica automatica, il modello sarà basato sulla prima variabile dell'elenco.
- Se si selezionano più variabili per la ricodifica e si sceglie l'opzione Usa lo stesso schema di ricodifica per tutte le variabili e/o si seleziona Applica modello, il modello contiene lo schema di ricodifica automatica combinato per tutte le variabili

Applica modello da. Applica un modello di ricodifica automatica già salvato alle variabili selezionate per la ricodifica ed aggiunge tutti i valori aggiuntivi trovati nelle variabili alla fine dello schema, mantenendo la relazione tra i valori originali e quelli ricodificati automaticamente contenuti nello schema salvato.

- Tutte le variabili selezionate per la ricodifica devono essere dello stesso tipo (numeriche o stringa) ed il tipo deve corrispondere a quello definito nel modello.
- I modelli non contengono informazioni sui valori mancanti definiti dall'utente. I valori mancanti definiti dall'utente delle variabili target sono basati sulla prima variabile dell'elenco delle variabili originali con valori mancanti definiti dall'utente. Tutti gli altri valori derivanti dalle variabili originali, ad eccezione di quelli mancanti definiti dal sistema, sono considerati validi.

- Le associazioni dei valori del modello vengono applicate per prime. Tutti i valori restanti vengono ricodificati in valori superiori all'ultimo valore del modello, mentre i valori mancanti definiti dall'utente (basati sulla prima variabile dell'elenco con valori mancanti definiti dall'utente) vengono ricodificati in valori superiori all'ultimo valore valido.
- Se si selezionano più variabili per la ricodifica automatica, viene applicato prima il modello, seguito da una ricodifica automatica combinata comune per tutti i valori aggiuntivi trovati nelle variabili selezionate. Ciò genera uno schema di ricodifica automatica singolo di tipo comune per tutte le variabili selezionate.

Per ricodificare stringhe o valori numerici in interi consecutivi

▶ Dai menu, scegliere:
Trasforma
Ricodifica automatica...

▶ Selezionare una o più variabili da ricodificare.

▶ Per ciascuna variabile selezionata, specificare un nuovo nome di variabile e fare clic su Nuovo nome.

Classifica casi

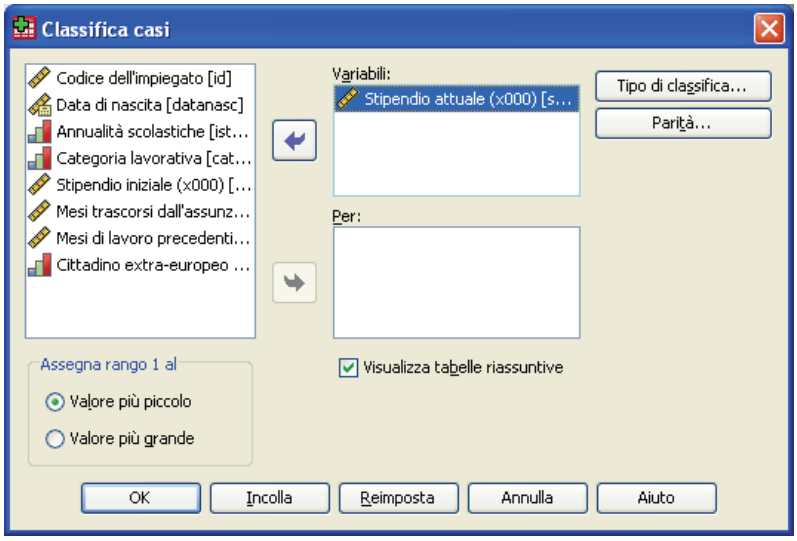
La finestra di dialogo Classifica casi permette di creare nuove variabili contenenti ranghi, punteggi normali ed esponenziali e valori percentili per le variabili numeriche.

I nomi delle nuove variabili e le etichette delle variabili descrittive verranno generate automaticamente in base al nome originale della variabile e alle misure selezionate. Nella tabella riassuntiva verranno elencate le variabili originali, le nuove variabili e le etichette delle variabili.

Se lo si desidera, è possibile:

- Classificare i casi in ordine crescente o decrescente.
- Organizzare ranghi in sottogruppi selezionando una o più variabili di raggruppamento per la lista Per. Verranno creati ranghi all'interno di ciascun gruppo. I gruppi vengono definiti dalla combinazione dei valori delle variabili di raggruppamento. Ad esempio, se si seleziona *sesso* e *minoranza* come variabili di raggruppamento, verranno creati ranghi per ciascuna combinazione di *sesso* e *minoranza*.

Figura 8-13
Finestra di dialogo Classifica casi



Per classificare i casi

▶ Dai menu, scegliere:
Trasforma
Classifica casi...

▶ Selezionare una o più variabili da classificare. È possibile classificare solo variabili numeriche.

È possibile classificare i casi in ordine crescente o decrescente e organizzarli in sottogruppi.

Classifica casi: Tipi

È possibile selezionare più metodi di classifica. Per ciascun metodo verrà creata una variabile di rango distinta. Sono disponibili diversi metodi di classifica, ovvero rango semplice, punteggi esponenziali, punteggi frazionari e percentili. È inoltre possibile creare classifiche in base alle stime di proporzione e ai punteggi normali.

Rango. Il rango semplice, espresso come intero positivo. Il valore della nuova variabile corrisponde al rango.

Punteggio esponenziale. La nuova variabile contiene punteggi esponenziali basati su una distribuzione esponenziale.

Rango frazionario. Il valore della nuova variabile è uguale al rango diviso per la somma dei pesi dei casi non mancanti.

Rango frazionario come percentuale. Ogni rango è diviso per il numero di casi con valori validi e moltiplicato per 100.

Somma dei pesi di caso. Il valore della nuova variabile è uguale alla somma dei pesi di caso. La nuova variabile è una costante per tutti i casi nello stesso gruppo.

N percentili. I ranghi sono definiti in base a gruppi percentili, dove ogni gruppo contiene approssimativamente lo stesso numero di casi. Ad esempio, 4 percentili assegnerebbe rango 1 ai casi compresi entro il 25° percentile, 2 ai casi compresi fra il 25° e il 50° percentile, 3 ai casi compresi fra il 50° e il 75° percentile, 4 a quelli oltre il 75°.

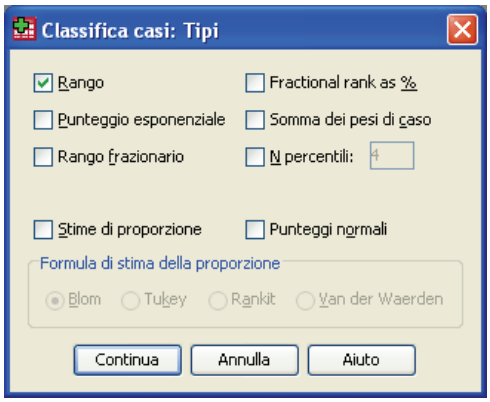
Stime di proporzione. Sono stime della proporzione cumulata di distribuzione in corrispondenza di un rango.

Punteggi normali. I punteggi normali sono i punteggi Z corrispondenti alla proporzione cumulata stimata.

Formula di stima della proporzione. Per le stime di proporzione e i punteggi normali è possibile selezionare la formula di stima della proporzione: Blom, Tukey, Rankit o Van der Waerden.

- **Blom.** Crea una nuova classifica basata su stime di proporzione. Usa la formula (r-3/8) / (w+1/4), dove r è il rango e w è la somma dei pesi dei casi.
- **Tukey.** Usa la formula (r-1/3) / (p+1/3), dove r è il rango e p è la somma dei pesi.
- **Rankit.** Usa la formula (r-1/2) / w, dove w è il numero di osservazioni e r è il rango, che varia fra 1 e w.
- **Van der Waerden.** Trasformazione di Van der Waerden, definita dalla formula r/(p+1), dove p è la somma dei pesi e r è il rango, che varia da 1 a p.

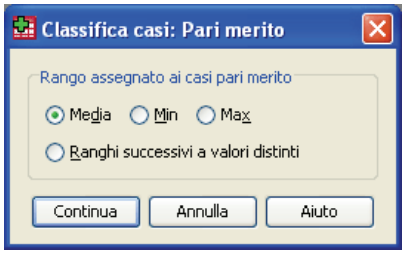
Figura 8-14
Finestra di dialogo Classifica casi: Tipi



Classifica casi: Pari merito

Questa finestra di dialogo consente di controllare il metodo di assegnazione dei ranghi ai casi con lo stesso valore della variabile originale.

Figura 8-15
Finestra di dialogo Classifica casi: Pari merito



Nella tabella che segue vengono indicati i diversi metodi di assegnazione dei ranghi ai valori a pari merito.

Valore	Media	Min	Max	Sequenziale
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

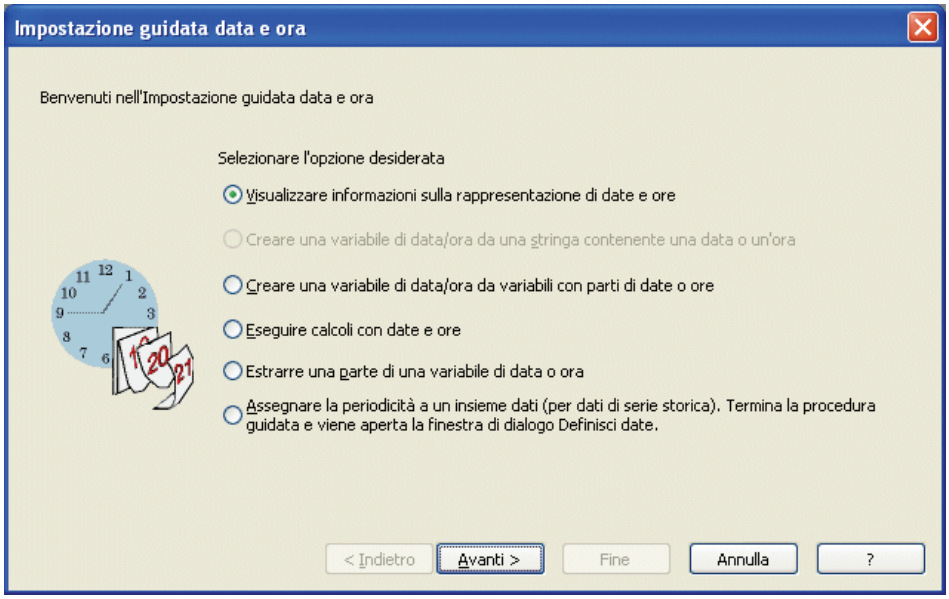
Impostazione guidata data e ora

L'Impostazione guidata data/ora semplifica molte operazioni normalmente associate alle variabili data/ora.

Per utilizzare l'Impostazione guidata data e ora

- Dai menu, scegliere:
  Trasforma
  Procedura guidata Data e ora...
- Selezionare l'operazione da eseguire e seguire le istruzioni per definire l'operazione.

Figura 8-16
Schermata introduttiva dell'Impostazione guidata data e ora


Impostazione guidata data e ora
Benvenuti nell'Impostazione guidata data e ora
Selezionare l'opzione desiderata
Visualizzare informazioni sulla rappresentazione di date e ore
Creare una variabile di data/ora da una stringa contenente una data o un'ora
Creare una variabile di data/ora da variabili con parti di date o ore
Eseguire calcoli con date e ore
Estrarre una parte di una variabile di data o ora
Assegnare la periodicità a un insieme dati (per dati di serie storica). Termina la procedura guidata e viene aperta la finestra di dialogo Definisci date.
< Indietro
Avanti >
Fine
Annulla
?

- **Visualizzare informazioni sulla rappresentazione di date e ore.** Questa opzione visualizza una schermata che fornisce una breve panoramica sulle variabili data/ora di SPSS Statistics. Se si seleziona il pulsante ?, viene visualizzato anche un collegamento con informazioni più dettagliate.
- **Creazione di una variabile di data/ora da una variabile stringa che contiene una data o un orario.** Utilizzare questa opzione per creare una variabile data/ora da una variabile stringa. Ad esempio se la variabile stringa utilizza rappresenta le date nel formato mm/gg/aaaa e si desidera utilizzarla per creare una variabile data/ora.
- **Crea una variabile data/ora dalle variabili che contengono alcuni elementi di date o ore.** Questa opzione permette di creare una variabile data/ora da un insieme di variabili esistenti. Esempio: si supponga di disporre di una variabile che rappresenta il mese (sotto forma di intero), di una seconda variabile che rappresenta il giorno del mese e di una terza variabile che rappresenta l'anno. Le tre variabili possono essere combinate in un'unica variabile data/ora.
- **Eseguire calcoli con date e ore.** Utilizzare questa opzione per aggiungere o sottrarre valori dalle variabili data/ora. Ad esempio è possibile calcolare la durata di un processo sottraendo la variabile che rappresenta l'ora di inizio del processo dalla variabile che rappresenta l'ora di fine.
- **Estrai una parte della variabile data/ora.** Questa opzione permette di estrarre una parte di una variabile di data o ora; ad esempio il giorno del mese da una variabile di data/ora con formato mm/gg/aaaa.
- **Assegna periodicità a file di dati.** Questa opzione consente di visualizzare la finestra di dialogo Definisci date, usata per creare le variabili data/ora costituite da un insieme di date in sequenza. Questa opzione viene generalmente usata per associare date ai dati con serie temporali.

Nota: le attività sono disattivate quando il file di dati non contiene i tipi di variabili richiesti per effettuare l'attività. Ad esempio, se il file di dati contiene variabili non stringa, l'attività di creazione di una variabile data/ora da una stringa non è applicabile ed è disattivata.

Date e ore in SPSS Statistics

Le variabili che rappresentano le date e le ore in SPSS Statistics sono variabili di tipo numerico, che visualizzano formati che corrispondono ai formati di data/ora specifici. Queste variabili vengono generalmente chiamate variabili data/ora. Viene fatta una distinzione tra le variabili data/ora che rappresentano le date e quelle che rappresentano una durata indipendente dalle date; ad esempio 20 ore, 10 minuti e 15 secondi. Queste ultime sono esempi di variabili di durata e si distinguono dalle prime che sono variabili data/ora. Per un elenco completo dei formati di visualizzazione, vedere "Data e ora" nella sezione "Formati universali" di *Command Syntax Reference*.

Variabili data e data/ora. Le variabili data hanno un formato che rappresenta una data; ad esempio mm/gg/aaaa. Le variabili data/ora hanno un formato che rappresenta una data ed un'ora; ad gg-mm-aaaa e hh:mm:ss. Internamente, le variabili data e data/ora vengono salvate in numero di secondi calcolati a partire dal 14 ottobre 1582. Le variabili di data e data/ora sono talvolta chiamate variabile del formato data.

- È possibile specificare l'anno sia con due che con quattro cifre. Per impostazione predefinita, gli anni a due cifre comprendono un intervallo che inizia 69 anni prima e termina 30 anni dopo la data corrente. Questo intervallo varia a seconda delle impostazioni selezionate per Opzioni e può essere configurato selezionando Opzioni e facendo clic su Dati nel menu Modifica.
- I trattini, i punti, le virgole, le barre o gli spazi possono essere usati come delimitatori nei formati giorno-mese-anno.
- I mesi possono essere rappresentati in cifre, numeri romani, abbreviazioni di tre lettere o espressi con il nome completo. Le abbreviazioni di tre lettere o il nome completo devono essere scritti in inglese. I nomi di mesi in altre lingue non vengono riconosciuti.

Variabili durata. Le variabili durata hanno un formato che rappresenta un intervallo di tempo; ad esempio hh:mm. Queste variabili vengono salvate internamente come secondi senza riferimento a una data specifica.

- Nelle specifiche di tempo (applicabili a variabili data/ora e di durata), è possibile usare il punto e virgola per delimitare le ore, i minuti e i secondi. Le ore e i minuti devono sempre essere specificati; i secondi sono opzionali. È sempre necessario inserire un punto per separare i secondi dalle frazioni di secondo. Le ore possono avere qualsiasi grandezza, ma il valore massimo è 59 per i minuti e 59,999.... per i secondi.

Data e ora correnti. La variabile di sistema *$TIME* contiene la data e l'ora correnti. La variabile rappresenta il numero di secondi calcolati dal1 4 ottobre 1582 alla data e all'ora in cui è stato eseguito il comando di trasformazione.

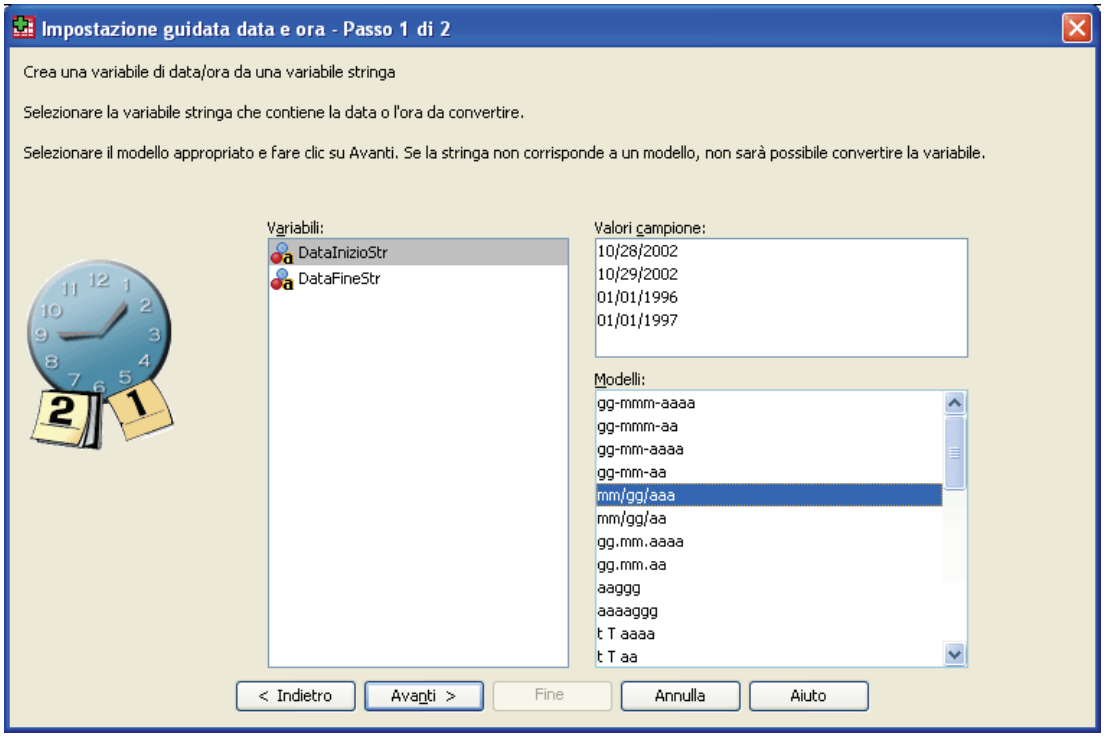
Creazione di una variabile data/ora da una stringa

Per creare una variabile data/ora da una variabile stringa:

▶ Selezionare Crea una variabile data/ora da una stringa contenente una data o un'ora nella pagina iniziale dell'Impostazione guidata data/ora.

Selezione della variabile stringa da convertire in una variabile data/ora

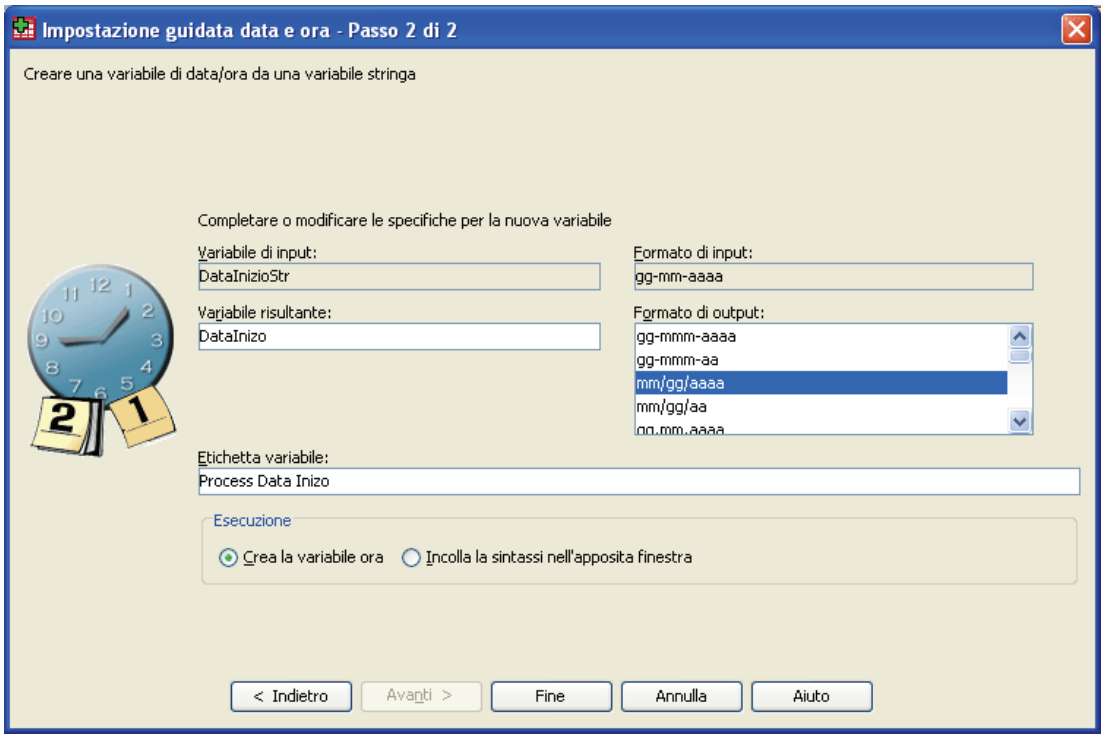
Figura 8-17
Creazione di una variabile data/ora dalla variabile stringa - Passaggio 1



- Selezionare la variabile stringa da convertire nell'elenco Variabili. È utile ricordare che l'elenco comprende solo variabili stringa.
- Selezionare un motivo dall'elenco Motivo, verificando che corrisponda alla modalità di rappresentazione delle date nella variabile stringa. L'elenco Valori campione visualizza i valori effettivi della variabile selezionata nel file di dati. I valori della variabile stringa che non corrispondono al motivo selezionato generano un valore mancante di sistema per la nuova variabile.

Selezione del risultato della conversione della variabile stringa in una variabile data/ora

Figura 8-18
Creazione di una variabile data/ora dalla variabile stringa - Passaggio 2



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.

  Se lo si desidera, è possibile:

  - Selezionare un formato data/ora per la nuova variabile nell'elenco Formato dell'output.
  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

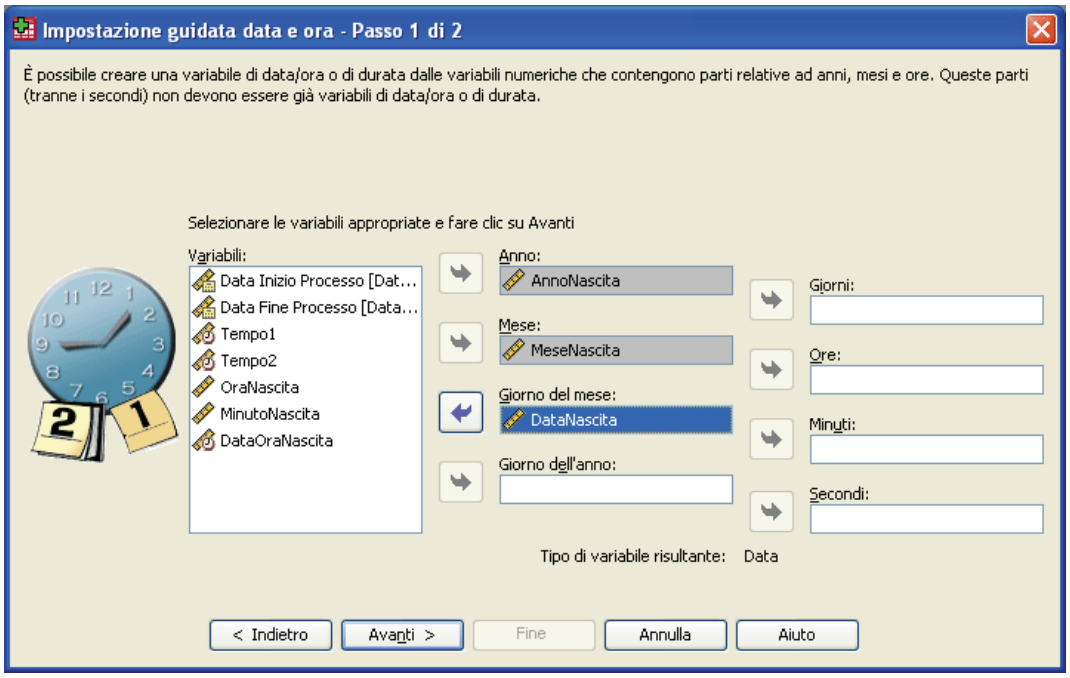
Creazione di una variabile data/ora da un insieme di variabili

Per unire un insieme di variabili esistenti un un'unica variabile data/ora:

- Selezionare Crea una variabile data/ora dalle variabili che contengono parti di date o ore nella pagina iniziale dell'Impostazione guidata data/ora.

Selezione delle variabili da unire alla variabile data/ora

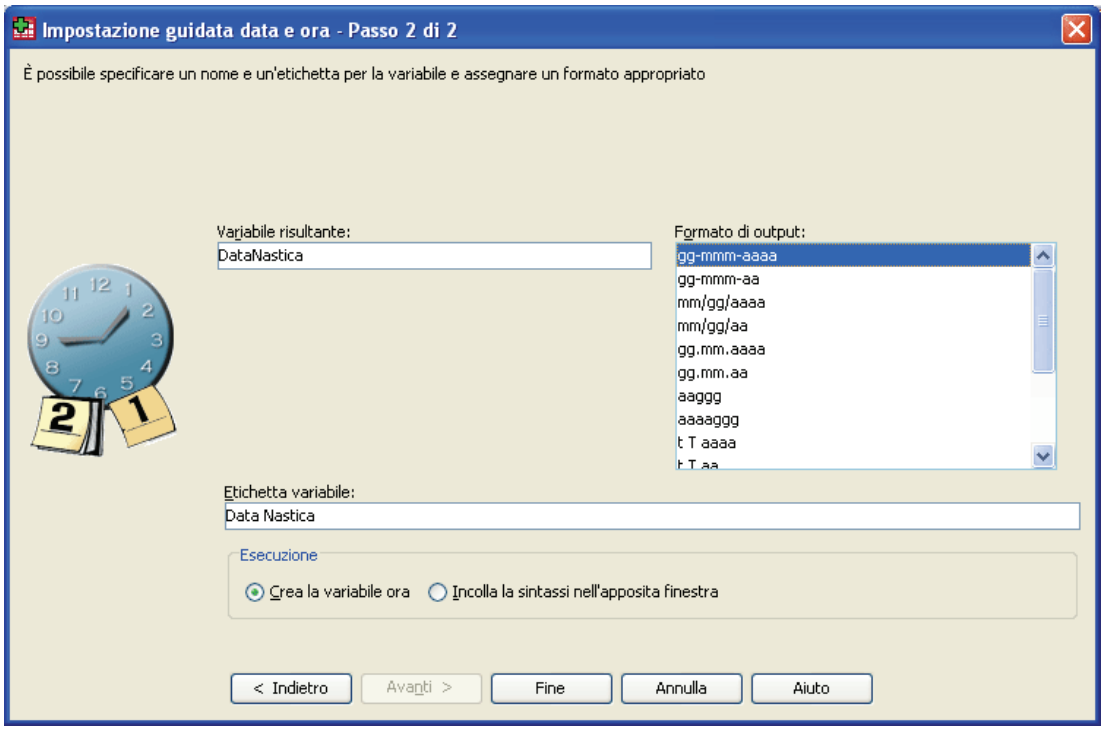
Figura 8-19
Creazione di una variabile data/ora dall'insieme di variabili - Passaggio 1



- Selezionare le variabili che rappresentano le diverse parti della data/dell'ora.
  - Ricordare che non è possibile usare tutte le combinazioni disponibili. Ad esempio non è possibile creare una variabile data/ora da Anno e Giorno del mese perché quando si seleziona Anno è necessario specificare un data completa.
  - Non è possibile usare una variabile data/ora esistente come parte della variabile data/ora finale che si sta creando. Le variabili che costituiscono le parti della nuova variabile data/ora devono essere numeri interi. L'unica eccezione è rappresentata dalla possibilità di usare una variabile data/ora esistente per la parte Secondi della nuova variabile. Poiché è possibile usare le frazioni di secondi, la variabile usata per i Secondi non deve necessariamente essere un numero intero.
  - I valori per qualsiasi parte della nuova variabile che non rientrano nel risultato consentito generano una variabile mancante di sistema per la nuova variabile. Ad esempio, se inavvertitamente è stata usata una variabile che rappresenta il giorno del mese per Mese, a qualsiasi caso con i valori del giorno del mese compresi nell'intervallo 14–31 verrà assegnato il valore mancante di sistema per la nuova variabile dal momento che l'intervallo valido per i mesi in SPSS Statistics è 1–13.

Selezione di una variabile data/ora creata mediante l'unione di variabili

Figura 8-20
Creazione di una variabile data/ora dall'insieme di variabili - Passaggio 2



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.
- Selezionare un formato di data/ora dall'elenco Formato dell'output.

  Se lo si desidera, è possibile:

  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

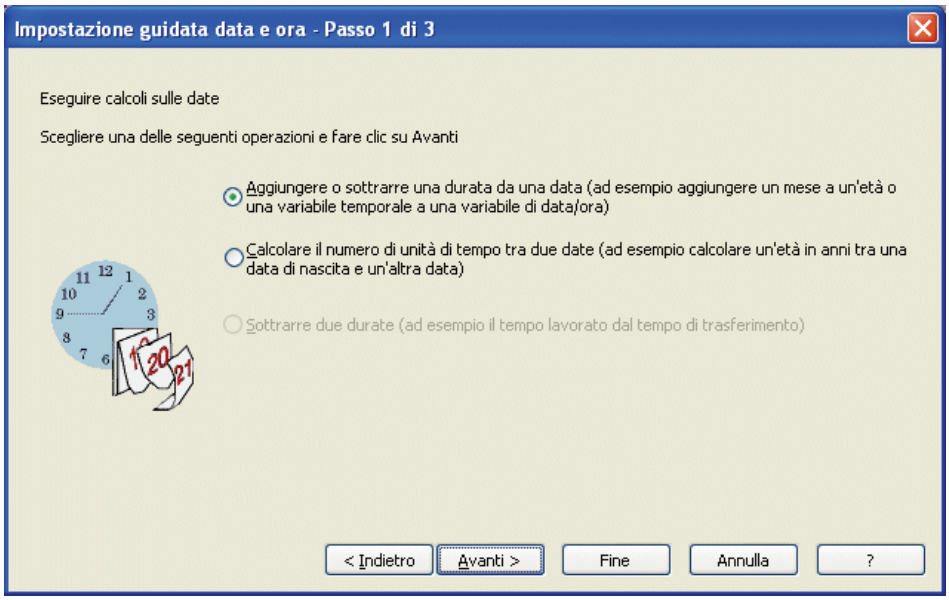
Aggiunta o sottrazione di valori dalle variabili data/ora

Per aggiungere o sottrare valori dalle variabili data/ora:

- Selezionare Eseguire calcoli con date e ore nella pagina iniziale dell'Impostazione guidata data/ora.

Selezione del tipo di calcolo da eseguire con le variabili data/ora

Figura 8-21
Aggiunta o sottrazione di valori alle/dalle variabili data/ora - Passaggio 1


Impostazione guidata data e ora - Passo 1 di 3
Eseguire calcoli sulle date
Scegliere una delle seguenti operazioni e fare clic su Avanti
Aggiungere o sottrarre una durata da una data (ad esempio aggiungere un mese a un'età o una variabile temporale a una variabile di data/ora)
Calcolare il numero di unità di tempo tra due date (ad esempio calcolare un'età in anni tra una data di nascita e un'altra data)
Sottrarre due durate (ad esempio il tempo lavorato dal tempo di trasferimento)
< Indietro
Avanti >
Fine
Annulla
?

- **Aggiungere o sottrarre una durata da una data.** Utilizzare questa opzione per aggiungere o sottrarre a una/da una variabile formato data. È possibile aggiungere o sottrarre durate corrispondenti a valori fissi come 10 giorni oppure valori da una variabile numerica; ad esempio una variabile che rappresenta gli anni.
- **Calcolare il numero di unità di tempo tra due date.** Usare questa opzione per ottenere la differenza tra due date nell'unità selezionata. Ad esempio è possibile ottenere il numero di anni o il numero di giorni che separano due date.
- **Sottrarre due durate.** Usare questa opzione per ottenere la differenza tra due variabili che hanno formati di durata; ad esempio hh:mm o hh:mm:ss.

Nota: le attività sono disattivate quando il file di dati non contiene i tipi di variabili richiesti per effettuare l'attività. Ad esempio se il file di dati non contiene due variabili con formati di durata, l'operazione di sottrazione delle due durate non è applicabile e viene disattivata.

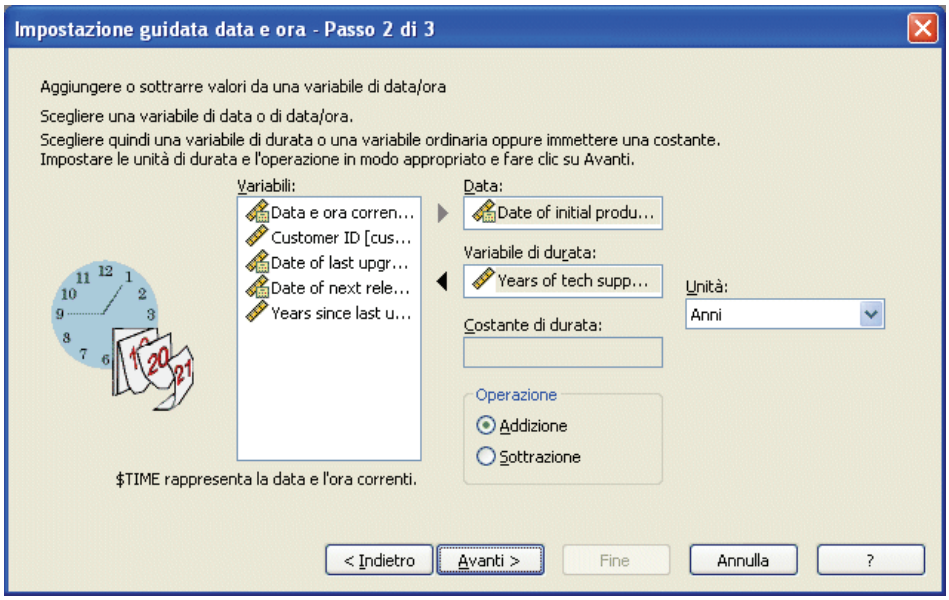
Aggiunta o sottrazione di una durata a/da una data

Per aggiungere o sottrarre una durata da una variabile formato data:

▶ Selezionare Aggiungi o sottrai una durata alla/dalla data nella schermata *Esegui calcoli sulle date* dell'Impostazione guidata data/ora.

Selezione della variabile data/ora e della durata da aggiungere o sottrarre

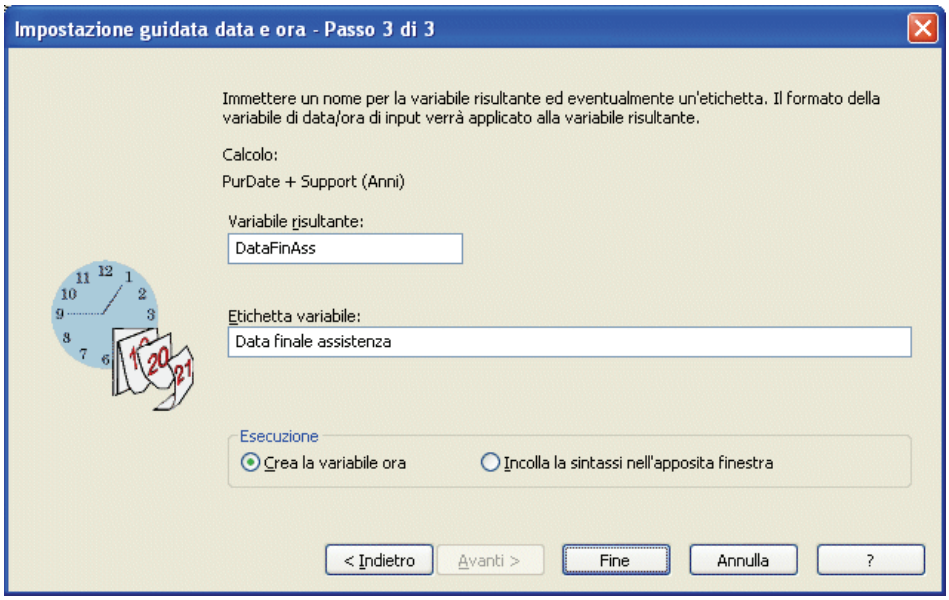
Figura 8-22
Aggiunta o sottrazione della durata - Passaggio 2


Impostazione guidata data e ora - Passo 2 di 3
Aggiungere o sottrarre valori da una variabile di data/ora
Scegliere una variabile di data o di data/ora.
Scegliere quindi una variabile di durata o una variabile ordinaria oppure immettere una costante.
Impostare le unità di durata e l'operazione in modo appropriato e fare clic su Avanti.
Variabili:
Data e ora corren...
Customer ID [cus...
Date of last upgr...
Date of next rele...
Years since last u...
Data:
Date of initial produ...
Variabile di durata:
Years of tech supp...
Unità:
Anni
Costante di durata:
Operazione
Addizione
Sottrazione
$TIME rappresenta la data e l'ora correnti.
< Indietro
Avanti >
Fine
Annulla
?

- Selezionare una variabile data (o ora).
- Selezionare una variabile durata oppure immettere un valore in Costante durata. Le variabili usate per le durate non possono essere variabili data o data/ora. Tuttavia, possono essere variabili di durata o semplici variabili numeriche.
- Selezionare l'unità della durata dall'elenco a discesa. Selezionare Durata se si utilizza una variabile di durata come hh:mm o hh:mm:ss.

Selezione del risultato dell'aggiunta o sottrazione di una durata a/da una variabile data/ora

Figura 8-23
Aggiunta o sottrazione della durata - Passaggio 3


Impostazione guidata data e ora - Passo 3 di 3
Immettere un nome per la variabile risultante ed eventualmente un'etichetta. Il formato della variabile di data/ora di input verrà applicato alla variabile risultante.
Calcolo:
PurDate + Support (Anni)
Variabile risultante:
DataFinAss
Etichetta variabile:
Data finale assistenza
Esecuzione
Crea la variabile ora
Incolla la sintassi nell'apposita finestra
< Indietro
Avanti >
Fine
Annulla
?

- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.

  Se lo si desidera, è possibile:

  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

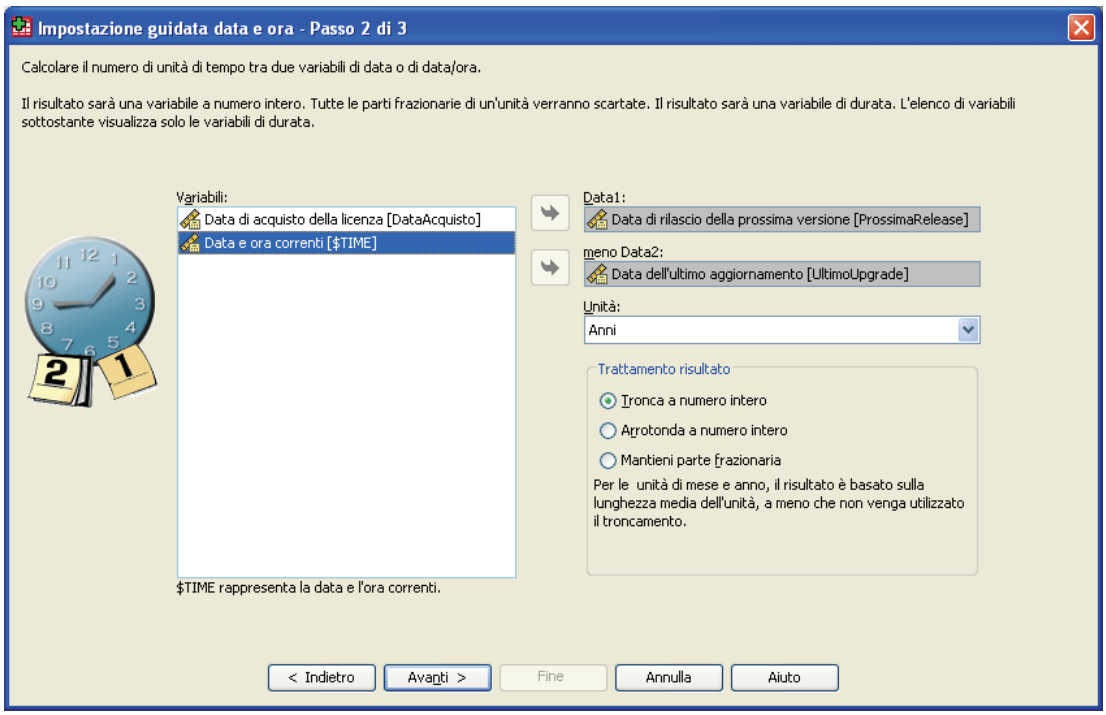
Sottrazione di variabili in formato data

Per sottrarre due variabili in formato data:

- Selezionare Calcola numero di unità di tempo tra due date nella schermata *Esegui calcoli sulle date* dell'Impostazione guidata data/ora.

Selezione delle variabli in formato data da sottrarre

Figura 8-24
Sottrazione di date - Passaggio 2


Impostazione guidata data e ora - Passo 2 di 3
Calcolare il numero di unità di tempo tra due variabili di data o di data/ora.
Il risultato sarà una variabile a numero intero. Tutte le parti frazionarie di un'unità verranno scartate. Il risultato sarà una variabile di durata. L'elenco di variabili sottostante visualizza solo le variabili di durata.
Variabili:
Data di acquisto della licenza [DataAcquisto]
Data e ora correnti [$TIME]
$TIME rappresenta la data e l'ora correnti.
Data1:
Data di rilascio della prossima versione [ProssimaRelease]
meno Data2:
Data dell'ultimo aggiornamento [UltimoUpgrade]
Unità:
Anni
Trattamento risultato
Tronca a numero intero
Arrotonda a numero intero
Mantieni parte frazionaria
Per le unità di mese e anno, il risultato è basato sulla lunghezza media dell'unità, a meno che non venga utilizzato il troncamento.
< Indietro
Avanti >
Fine
Annulla
Aiuto

- Selezionare le variabili da sottrarre.
- Selezionare l'unità per il risultato dall'elenco a discesa.
- Selezionare come deve essere calcolato il risultato (Trattamento risultato).

Trattamento risultato

Le opzioni seguenti sono disponibili per la modalità di calcolo del risultato:

- **Tronca a numero intero.** Viene ignorata qualsiasi parte frazionaria del risultato. Ad esempio, sottraendo 10/28/2006 da 10/21/2007 viene restituito un risultato di 0 per gli anni e 11 per i mesi.
- **Arrotonda a numero intero.** Il risultato viene arrotondato al numero intero più vicino. Ad esempio, sottraendo 10/28/2006 da 10/21/2007 viene restituito un risultato di 1 per gli anni e 12 per i mesi.
- **Mantieni parte frazionaria.** Viene mantenuto il valore completo; non viene applicato nessun arrotondamento o troncamento. Ad esempio, sottraendo 10/28/2006 da 10/21/2007 viene restituito un risultato di 0,98 per gli anni e 11,76 per i mesi.

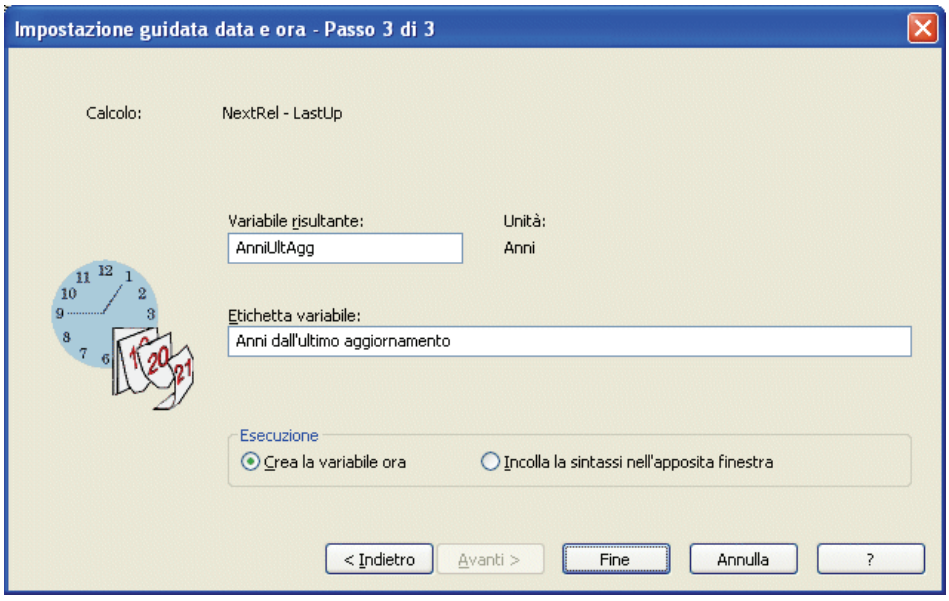
Per l'arrotondamento e il mantenimento della parte frazionaria, il risultato per gli anni è basato sul numero medio di giorni in un anno (365,25) e il risultato per i mesi è basato sul numero medio di giorni in un mese (30,4375). Ad esempio, sottraendo 2/1/2007 da 3/1/2007 (formato m/g/a) viene restituito un risultato frazionario di 0,92 mesi, mentre sottraendo 3/1/2007 da 2/1/2007 viene

restituita una differenza frazionaria di 1,02 mesi. Questo influisce anche sui valori calcolati in periodi di tempo che includono anni bisestili. Ad esempio, sottraendo 2/1/2008 da 3/1/2008 viene restituita una differenza frazionaria di 0,95 mesi, confrontata con 0,92 per lo stesso periodo di tempo in un anno non bisestile.

Data 1	Data 2	Anni			Mesi		
		Troncato	Arroton-dato	Frazione	Troncato	Arroton-dato	Frazione
10/21/2006	10/28/2007	1	1	1.02	12	12	12.22
10/28/2006	10/21/2007	0	1	.98	11	12	11.76
2/1/2007	3/1/2007	0	0	.08	1	1	.92
2/1/2008	3/1/2008	0	0	.08	1	1	.95
3/1/2007	4/1/2007	0	0	.08	1	1	1.02
4/1/2007	5/1/2007	0	0	.08	1	1	.99

Selezione del risultato per la sottrazione di due variabili in formato data

Figura 8-25
Sottrazione di date - Passaggio 3



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.

Se lo si desidera, è possibile:

- Assegnare un'etichetta variabile descrittiva alla nuova variabile.

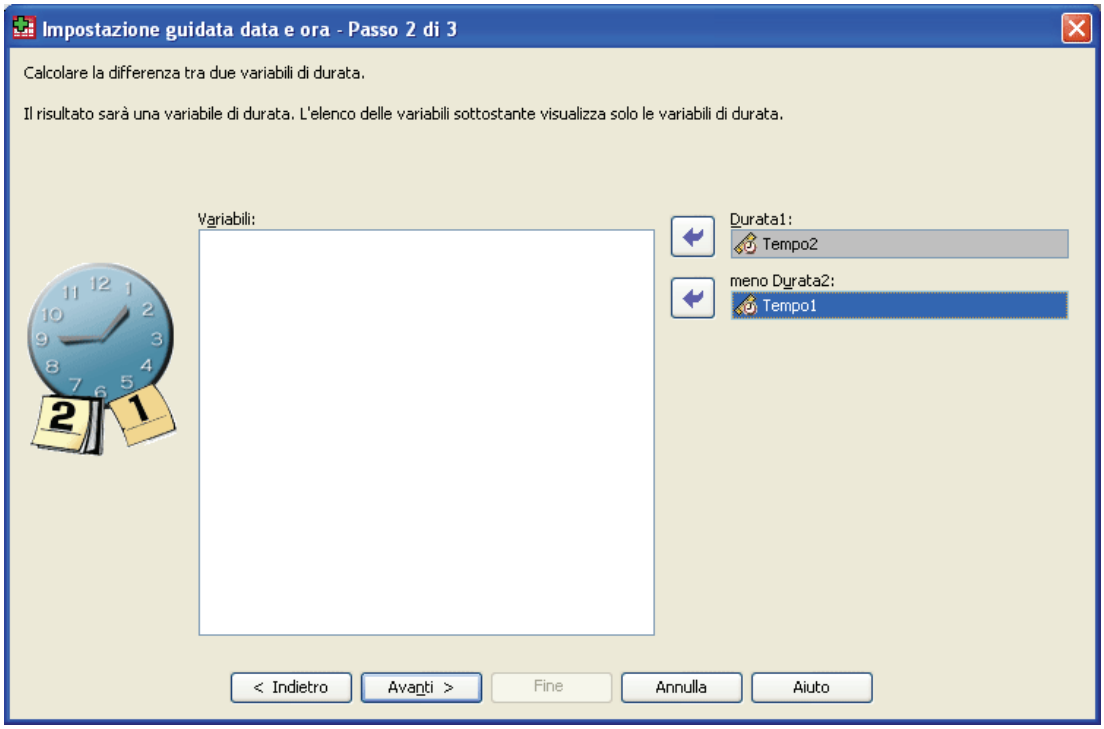
Sottrazione di variabili di durata

Per sottrarre due variabili di durata:

- Selezionare Sottrai due durate nella schermata *Esegui calcoli sulle date* dell'Impostazione guidata data/ora.

Selezione delle variabili di durata da sottrarre

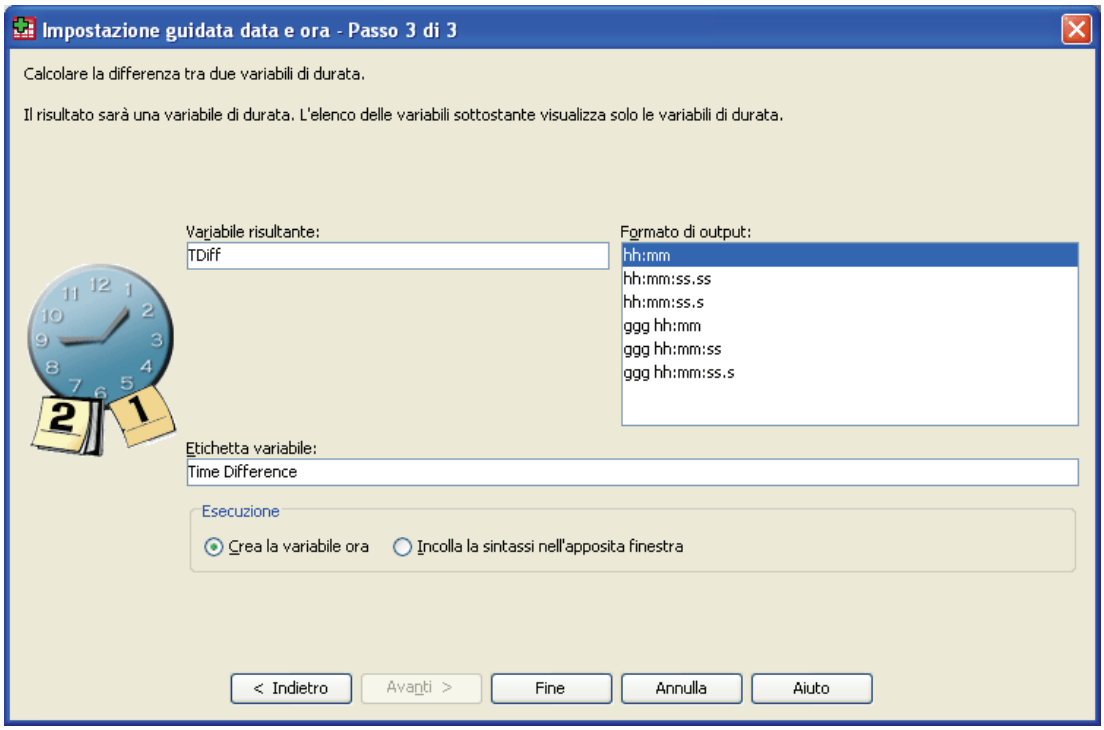
Figura 8-26
Sottrazione di due durate - Passo 2


Impostazione guidata data e ora - Passo 2 di 3
Calcolare la differenza tra due variabili di durata.
Il risultato sarà una variabile di durata. L'elenco delle variabili sottostante visualizza solo le variabili di durata.
Variabili:
Durata1:
Tempo2
meno Durata2:
Tempo1
< Indietro
Avanti >
Fine
Annulla
Aiuto

- Selezionare le variabili da sottrarre.

Selezione del risultato della sottrazione di due variabili di durata

Figura 8-27
Sottrazione di due durate - Passo 3



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.
- Selezionare un formato di durata dall'elenco Formato dell'output.

  Se lo si desidera, è possibile:

  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

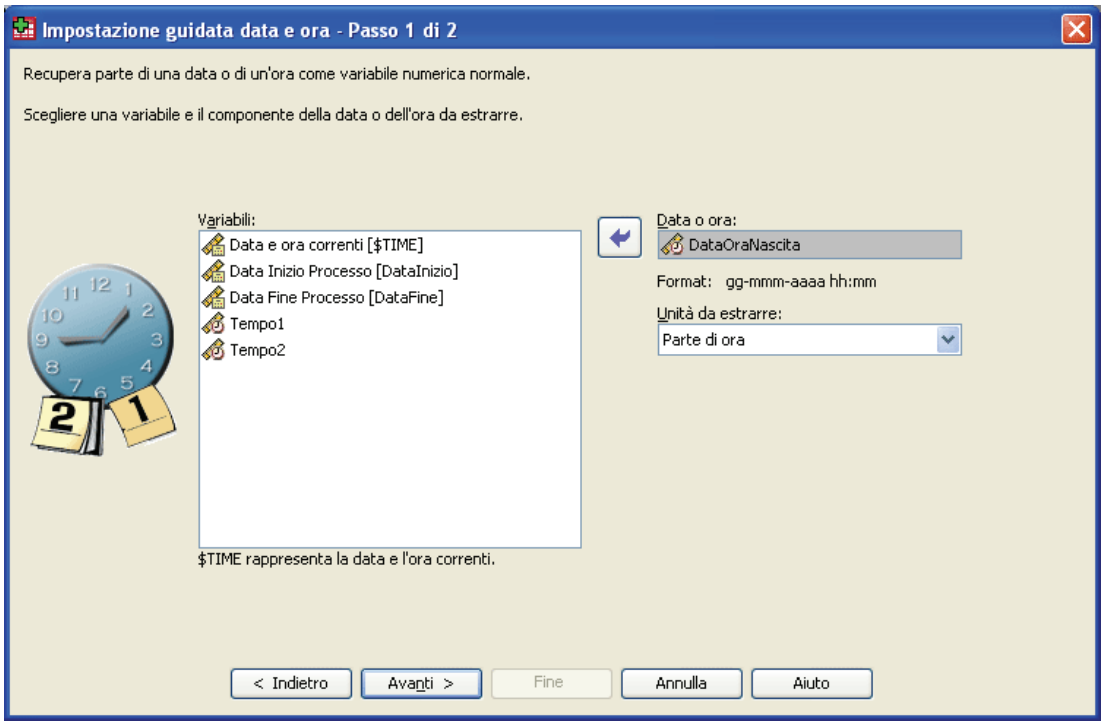
Estrazione di parte della variabile data/ora

Per estrarre un componente, come l'anno, da una variabile data/ora:

- Selezionare Estrai una parte della variabile data o ora nella schermata iniziale dell'Impostazione guidata data/ora.

Selezione del componente da estrarre dalla variabile data/ora

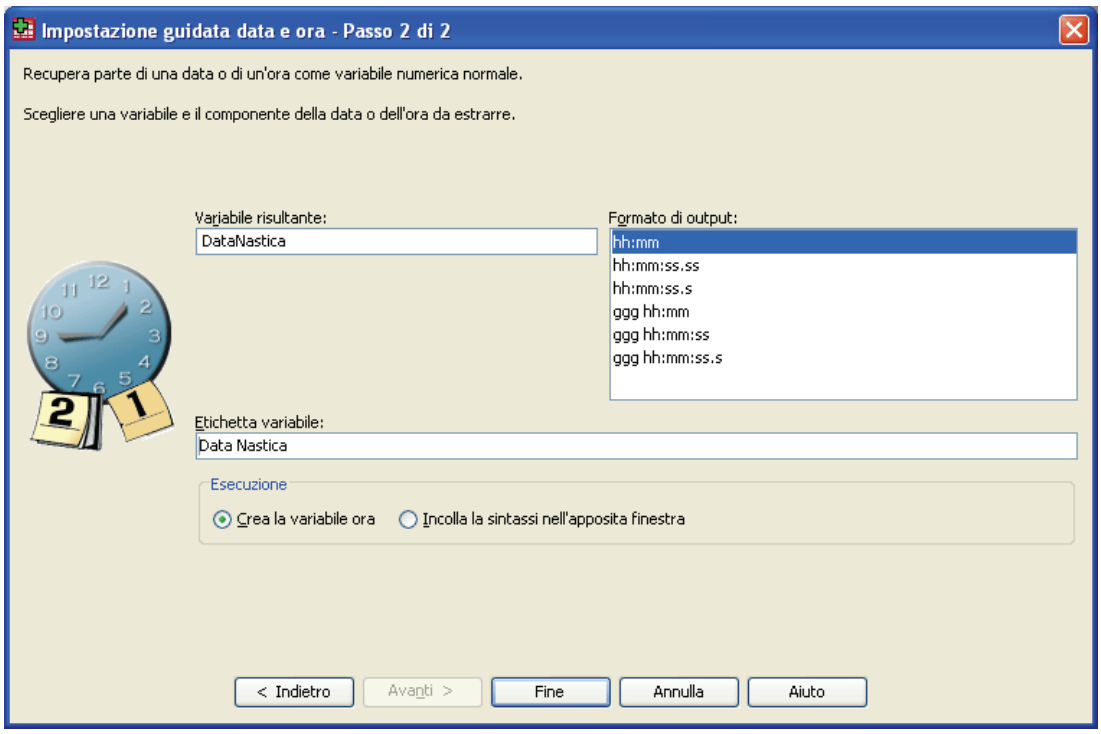
Figura 8-28
Acquisizione della parte della variabile data/ora - Passaggio 1


Impostazione guidata data e ora - Passo 1 di 2
Recupera parte di una data o di un'ora come variabile numerica normale.
Scegliere una variabile e il componente della data o dell'ora da estrarre.
Variabili:
Data e ora correnti [$TIME]
Data Inizio Processo [DataInizio]
Data Fine Processo [DataFine]
Tempo1
Tempo2
$TIME rappresenta la data e l'ora correnti.
Data o ora:
DataOraNascita
Format: gg-mmm-aaaa hh:mm
Unità da estrarre:
Parte di ora
< Indietro
Avanti >
Fine
Annulla
Aiuto

- Selezionare la variabile che contiene la parte di data o ora da estrarre.
- Selezionare la parte della variabile da estrarre dall'elenco a discesa. È possibile estrarre informazioni dalle date anche se non fanno esplicitamente parte della data visualizzata; come qualsiasi giorno della settimana.

Selezione del risultato per l'estrazione del componente dalla variabile data/ora

Figura 8-29
Acquisizione della parte della variabile data/ora - Passaggio 2



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.
- Se si sta estraendo la parte data o ora da una variabile data/ora, è necessario selezionare un formato dall'elenco Formato dell'output. Questo elenco è disattivato per i casi in cui non è richiesto.

  Se lo si desidera, è possibile:

  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

Trasformazioni di dati di serie storiche

Sono disponibili numerose trasformazioni di dati che risultano utili per l'analisi delle serie storiche:

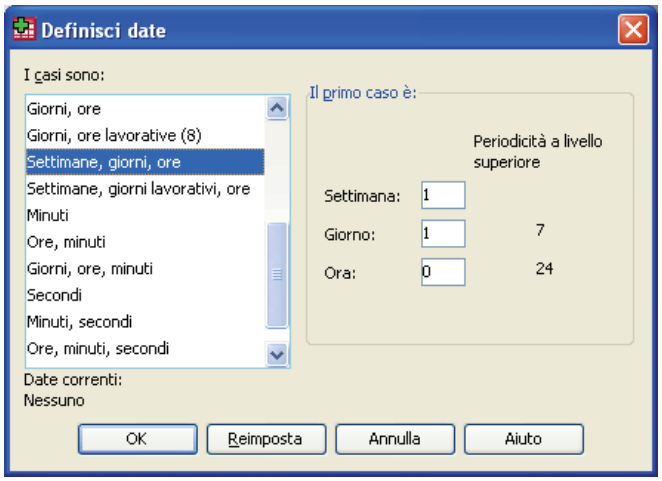
- Generazione di variabili della data per definire la periodicità e distinguere tra periodi storici, di convalida e di previsione.
- Creazione di nuove variabili di serie storica come funzioni di variabili di serie storica esistenti.
- Sostituzione di valori mancanti di sistema e definiti dall'utente con stime basate su diversi metodi.

Per ottenere una **serie storica** è necessario misurare una variabile (o un insieme di variabili) regolarmente per un periodo di tempo. Le trasformazioni di dati di serie storiche si fondano su una struttura del file di dati in cui ciascun caso (riga) rappresenta una serie di osservazioni eseguite in momenti diversi e con un periodo di tempo uniforme tra i casi.

Definisci date

La finestra di dialogo Definisci date consente di generare variabili della data che possono essere utilizzate per stabilire la periodicità di una **serie storica** ed applicare etichette all'output dell'analisi delle serie storiche.

Figura 8-30
Finestra di dialogo Definisci date



I casi sono. Consente di definire l'intervallo di tempo utilizzato per generare le date.

- Senza data. Consente di rimuovere le variabili della data precedentemente definite. Tutte le variabili con i seguenti nomi vengono eliminate: *year_*, *quarter_*, *month_*, *week_*, *day_*, *hour_*, *minute_*, *second_*, and *date_*.
- Personalizzato. Indica la presenza di variabili di data personalizzate create con la sintassi dei comandi (ad esempio, una settimana lavorativa di quattro giorni). Questo riflette semplicemente lo stato dell'insieme di dati attivo. Selezionandolo dall'elenco non si ottiene alcun risultato.

Il primo caso è. Consente di definire il valore iniziale che verrà assegnato al primo caso. Ai valori successivi verranno assegnati valori sequenziali in base all'intervallo di tempo.

Periodicità al livello più alto. Indica la variazione ciclica ripetitiva, ad esempio il numero di mesi dell'anno o il numero di giorni della settimana. Il valore visualizzato indica il massimo valore che è possibile inserire.

Per ciascun componente utilizzato per definire la data verrà creata una nuova variabile numerica. I nomi delle nuove variabili terminano con un carattere di sottolineatura. Per i componenti verrà inoltre creata una variabile stringa descrittiva, *date_*. Se, ad esempio, si seleziona Settimane, giorni, ore, verranno create quattro nuove variabili: *week_*, *day_*, *hour_* e *date_*.

Se i nomi delle variabili sono già stati definiti, verranno sostituiti quando si creano nuove variabili con gli stessi nomi delle variabili della data esistenti.

Per definire date per dati di serie storiche

- Dai menu, scegliere:
  Dati
  Definisci date...
- Selezionare un intervallo di tempo per l'elenco I casi sono.
- Specificare uno o più valori che definiscano la data iniziale per Il primo caso è, che determinerà la data assegnata al primo caso.

Variabili della data e variabili del formato della data

Le variabili della data create con l'opzione Definisci date non devono essere confuse con le variabili del formato della data, definite con l'opzione Visualizzazione variabili dell'Editor dei dati. Le variabili della data vengono utilizzate per stabilire la periodicità dei dati di serie storiche. Le variabili del formato della data rappresentano date e/o orari espressi in diversi formati. Le variabili della data sono interi semplici che rappresentano, ad esempio, il numero di giorni, settimane, ore a partire da un punto iniziale definito dall'utente. La maggior parte delle variabili del formato della data vengono memorizzate internamente come il numero di secondi trascorsi dal 14 ottobre 1582.

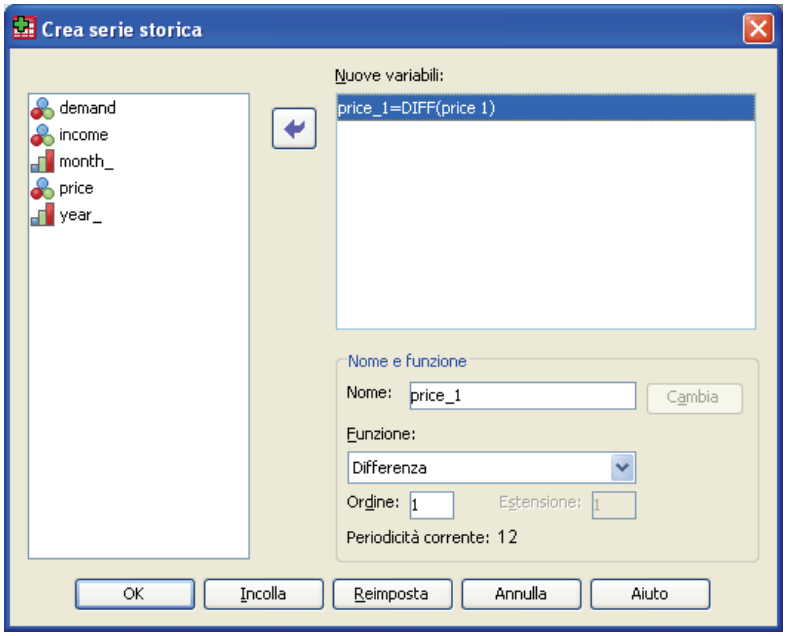
Crea serie storica

La finestra di dialogo Crea serie storica permette di creare nuove variabili basate su funzioni di variabili numeriche di **serie storica** esistenti. Questi valori trasformati risultano utili in numerose procedure di analisi delle serie storiche.

I nuovi nomi di variabile predefiniti sono rappresentati dai primi sei caratteri della variabile esistente in base alla quale sono stati creati, seguiti da un carattere di sottolineatura e da un numero sequenziale. Ad esempio, per la variabile *prezzo*, il nuovo nome di variabile sarà *prezzo_1*. Le nuove variabili mantengono le etichette dei valori definite per le variabili esistenti.

Per la creazione di variabili di serie storica sono disponibili diverse funzioni, ad esempio differenze, medie mobili, mediane mobili, ritardo e anticipo.

Figura 8-31
Finestra di dialogo Crea serie storica



Per creare nuove variabili di serie storica

- Dai menu, scegliere:
  Trasforma
  Crea serie storica...
- Selezionare la funzione di serie storica che si desidera utilizzare per trasformare le variabili originali.
- Selezionare le variabili in base alle quali creare nuove variabili di serie storica. È possibile utilizzare solo variabili numeriche.

  Se lo si desidera, è possibile:
  - Specificare nomi di variabili per annullare i nomi predefiniti delle nuove variabili.
  - Modificare la funzione relativa alla variabile selezionata.

Funzioni di trasformazione delle serie storiche

Differenza. Differenza non stagionale tra valori successivi della serie. L'ordine è rappresentato dal numero di valori precedenti utilizzati per calcolare la differenza. Poiché per ciascun ordine di differenza va perduta un'osservazione, i valori mancanti di sistema compaiono all'inizio della serie. Se, ad esempio, l'ordine della differenza è 2, ai primi due casi verrà assegnato il valore mancante di sistema per la nuova variabile.

Differenza stagionale. Differenza tra i valori della serie basata su un intervallo di tempo costante. L'intervallo viene definito in base alla periodicità corrente. Per calcolare le differenze stagionali, è innanzitutto necessario definire variabili della data (menu Dati, Definisci dati) che includano

un componente periodico (ad esempio i mesi dell'anno). L'ordine è rappresentato dal numero di periodi stagionali utilizzati per calcolare la differenza. Il numero di casi con il valore mancante di sistema all'inizio della serie equivale alla periodicità moltiplicata per l'ordine. Se, ad esempio, la periodicità corrente è 12 e l'ordine è 2, ai primi 24 casi verrà assegnato il valore mancante di sistema per la nuova variabile.

Media mobile centrata. Media di un intervallo di valori di serie che circondano e includono il valore corrente. L'intervallo è rappresentato dal numero di valori di serie utilizzati per calcolare la media. Se l'intervallo è pari, la media mobile verrà calcolata in base alla media di ciascuna coppia di medie non centrate. Il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie per un intervallo di *n* equivale a *n*/2 per gli intervalli di valori pari e (*n*–1)/2 per gli intervalli di valori dispari. Se, ad esempio, l'intervallo è 5, il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie è 2.

Media mobile sui precedenti. Media dell'intervallo di valori di serie che precedono il valore corrente. L'intervallo è rappresentato dal numero di valori di serie precedenti utilizzati per calcolare la media. Il numero di casi con il valore mancante di sistema all'inizio della serie equivale al valore dell'intervallo.

Mediane mobili. Mediane di intervalli di valori di serie che circondano e includono il valore corrente. L'intervallo è rappresentato dal numero di valori di serie utilizzati per calcolare la mediana. Se l'intervallo è pari, la mediana verrà calcolata in base alla media di ciascuna coppia di mediane non centrate. Il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie per un intervallo di *n* equivale a *n*/2 per gli intervalli di valori pari e (*n*–1)/2 per gli intervalli di valori dispari. Se, ad esempio, l'intervallo è 5, il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie è 2.

Somma cumulata. La somma cumulata dei valori di serie che precedono e includono il valore corrente.

Ritardo. Valore di un caso precedente, calcolato in base all'ordine di ritardo specificato. L'ordine è rappresentato dal numero di casi che precedono il valore corrente da cui si ottiene tale valore. Il numero di casi con il valore mancante di sistema all'inizio della serie equivale al valore dell'ordine.

Anticipo. Valore di un caso successivo, calcolato in base all'ordine di anticipo specificato. L'ordine è rappresentato dal numero di casi che seguono il caso corrente da cui si ottiene tale valore. Il numero di casi con il valore mancante di sistema alla fine della serie equivale al valore dell'ordine.

Livellamento. Nuovi valori di serie calcolati in base a un algoritmo di livellamento. Il livellatore inizia con la mediana mobile 4, centrata in base alla mediana mobile 2. I valori vengono quindi livellati nuovamente applicando la mediana mobile 5, la mediana mobile 3 e le medie mobili ponderate. I residui vengono calcolati sottraendo le serie livellate dalla serie originale. L'intero processo viene quindi ripetuto sui residui calcolati. Infine, vengono calcolati i residui livellati sottraendo i valori livellati ottenuti nella prima fase del processo. Questo procedimento talvolta viene definito **livellamento T4253H**.

Sostituisci valori mancanti

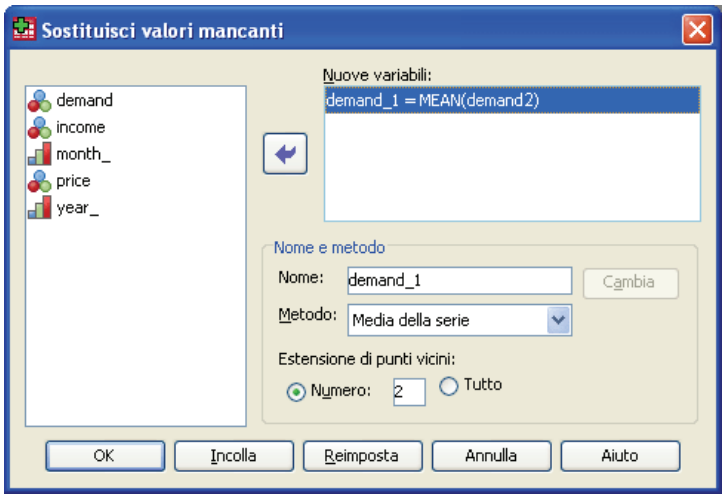
Le osservazioni mancanti possono creare problemi durante l'analisi e non è possibile calcolare alcune serie storiche misurate se queste includono valori mancanti. In alcuni casi il valore di un'osservazione specifica non è noto. Inoltre, i dati mancanti possono derivare da:

- Ciascun grado di differenziazione riduce la lunghezza della serie di 1.
- Ciascun grado di differenziazione stagionale riduce la lunghezza della serie di una stagione.
- Se si crea una nuova serie contenente previsioni che si estendono oltre la serie esistente (facendo clic su Salva e selezionando opzioni idonee), la serie originale e la serie residua risultante conterranno dati mancanti per le nuove osservazioni.
- Alcune trasformazioni come quella logaritmica generano dati mancanti per alcuni valori della serie originale.

I dati mancanti all'inizio o alla fine della serie non pongono problemi, perché riducono semplicemente la lunghezza utile della serie. Intervalli vuoti al centro di una serie (dati mancanti *incorporati*) possono porre problemi molto più seri. La portata del problema dipende dal tipo di procedura di analisi usato.

La finestra di dialogo Sostituisci valori mancanti consente di creare nuove variabili di **serie storica** da variabili esistenti, sostituendo i valori mancanti con le stime calcolate in base a uno dei metodi disponibili. I nuovi nomi di variabile predefiniti sono rappresentati dai primi sei caratteri della variabile esistente in base alla quale sono stati creati, seguiti da un carattere di sottolineatura e da un numero sequenziale. Ad esempio, per la variabile *prezzo*, il nuovo nome di variabile sarà *prezzo_1*. Le nuove variabili mantengono le etichette dei valori definite per le variabili esistenti.

Figura 8-32
Finestra di dialogo Sostituisci valori mancanti



Per sostituire i valori mancanti per le variabili di serie storica

▶ Dai menu, scegliere:

Trasforma
Sostituisci valori mancanti...

▶ Selezionare il metodo di stima che si desidera utilizzare per la sostituzione dei valori mancanti.

▶ Selezionare le variabili per cui si desidera sostituire i valori mancanti.

Se lo si desidera, è possibile:

- Specificare nomi di variabili per annullare i nomi predefiniti delle nuove variabili.
- Modificare il metodo di stima per la variabile selezionata.

Metodi di stima per la sostituzione dei valori mancanti

Media della serie. Consente di sostituire i valori mancanti con la media dell'intera serie.

Media di punti vicini. Consente di sostituire i valori mancanti con la media dei valori validi circostanti. L'intervallo di punti vicini è il numero di valori validi situati sopra e sotto il valore mancante utilizzato per calcolare la media.

Mediana di punti vicini. Consente di sostituire i valori mancanti con la mediana dei valori validi circostanti. L'intervallo dei punti vicini è il numero di valori validi situati sopra e sotto il valore mancante utilizzato per calcolare la mediana.

Interpolazione lineare. Consente di sostituire i valori mancanti utilizzando l'interpolazione lineare. Per l'interpolazione verranno utilizzati l'ultimo valore valido prima del valore mancante e il primo valore valido dopo il valore mancante. Se il primo o l'ultimo caso della serie contengono un valore mancante, quest'ultimo non verrà sostituito.

Trend lineare in quel punto. Consente di sostituire i valori mancanti con il trend lineare per quel punto. Sulla serie esistente verrà eseguita una regressione su una variabile indice che assume i valori da 1 a *n*. I valori mancanti verranno sostituiti con i rispettivi valori attesi.

Calcolo del punteggio dei dati con modelli predittivi

Il processo di applicare un modello predittivo ad un insieme di dati viene definito **calcolo del punteggio** dei dati. SPSS Statistics, Clementine e AnswerTree offrono procedure che consentono di creare modelli predittivi come i modelli di regressione, raggruppamento, albero e delle reti neurali. Dopo la creazione di un modello, le sue specifiche possono essere salvate come file XML che contiene tutte le informazioni necessarie per ricostruire il modello. Il prodotto SPSS Statistics Server fornisce il sistema per leggere il file del modello XML ed applicarlo al file di dati.

Esempio. Un'applicazione di credito viene utilizzata per valutare i rischi connessi alla concessione di un prestito. Il punteggio del credito che si ottiene dal modello di rischio può essere usato per decidere se concedere o meno il prestito.

Il calcolo del punteggio è equivalente ad una trasformazione dei dati. Il modello viene espresso internamente come un insieme di trasformazioni numeriche che possono essere applicate ad un insieme specifico di variabili - le variabili predittore del modello - per ottenere il risultato previsto. In questo senso il processo di calcolare il punteggio con un modello specifico è equivalente ad applicare una funzione, come una radice quadrata, ad un insieme di dati.

Il calcolo del punteggio è disponibile solo con SPSS Statistics Server e può essere eseguito interattivamente dagli utenti che utilizzano la modalità di analisi distribuita. Per calcolare il punteggio di file di dati di grandi dimensioni, è consigliabile ricorrere all'utilità Batch, una versione eseguibile separata fornita con SPSS Statistics Server. Per informazioni su come usare

l'utilità Batch Facility, vedere *Batch Facility User's Guide*, fornito come file PDF sul CD-ROM che contiene SPSS Statistics Server.

Il processo di calcolo del punteggio include quanto segue:

- Caricamento di un modello da un file in formato XML (PMML).
- Calcolo del punteggio come nuova variabile, utilizzando la funzione ApplyModel o StrApplyModel disponibile nella finestra di dialogo Calcola variabile.

  Per informazioni dettagliate sulla funzione `ApplyModel` o `StrApplyModel`, vedere le espressioni per il calcolo dei punteggi nella sezione Espressioni di trasformazione di *Command Syntax Reference*.

Nella tabella che segue sono elencate le procedure che supportano l'esportazione delle specifiche dei modelli in formato XML. È possibile utilizzare i modelli esportati con SPSS Statistics Server per calcolare il punteggio dei nuovi dati come descritto sopra. L'elenco completo dei tipi di modello di cui è possibile calcolare il punteggio con SPSS Statistics Server viene fornito nella descrizione della funzione `ApplyModel`.

Nome della procedura	**Nome del comando.**	**Opzione**
Discriminante	`DISCRIMINANT`	Base
Regressione lineare	`REGRESSIONE`	Base
Cluster TwoStep	`TWOSTEP CLUSTER`	Base
Generalized Linear Models	`GENLIN`	Advanced Statistics
Modello lineare generalizzato campioni complessi	`CSGLM`	Complex Samples
Regressione logistica campioni complessi	`CSLOGISTIC`	Complex Samples
Regressione ordinale campioni complessi	`CSORDINAL`	Complex Samples
Regressione logistica	`LOGISTIC REGRESSION`	Regression
Regressione logistica multinomiale	`NOMREG`	Regression
Albero decisionale	`TREE`	Decision Tree

Caricamento di un modello salvato

La finestra di dialogo Carica modello, che consente di caricare modelli predittivi salvati in formato XML (PMML), è disponibile solo quando si utilizza la modalità di analisi distribuita. Il caricamento dei modelli è il primo passo necessario per poter utilizzare tali modelli per il calcolo del punteggio dei dati.

Figura 8-33
Output di Carica modello

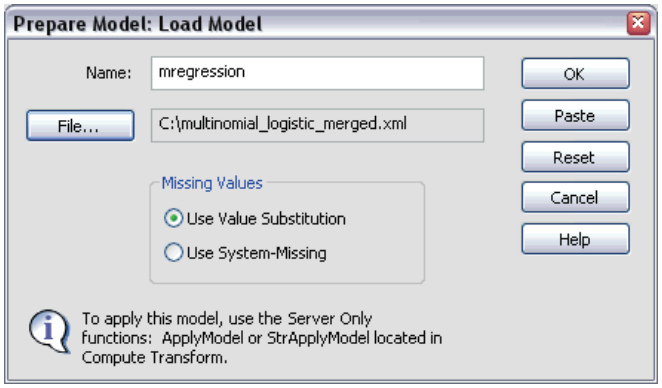
	Model Variables					
SPSS Variables	Name	Label	Type	Width	Role	Measurement
Age_Group	Age_Group	Age group	Numeric	8	Factor	Nominal
Frequency	Frequency	Number of purchases	Numeric	7	Covariate	Continuous
Gender	Gender	Gender	String	6	Factor	Nominal
Has_Broadband	Has_Broadband	Has home broadband connection	Numeric	10	Factor	Nominal
Has_Child	Has_Child	Has children	Numeric	10	Factor	Nominal
Income_Group	Income_Group	Income group	Numeric	10	Factor	Nominal
Log_Amount	Log_Amount	Log-transformed amount	Numeric	8	Covariate	Continuous
Recency	Recency	Days since last purchase	Numeric	10	Covariate	Continuous
Response	Response	Response	Numeric	10	Target	Nominal

Per caricare un modello

▶ Dai menu, scegliere:

Trasforma
  Prepara modello
    Carica modello...

Figura 8-34
Finestra di dialogo Carica modello: Prepara modello



▶ Specificare un nome da associare a questo modello. A ogni modello caricato è necessario assegnare un nome univoco.

▶ Fare clic su File e selezionare un file del modello. La finestra di dialogo Apri file che viene visualizzata indica i file disponibili in modalità di analisi distribuita. Sono inclusi i file presenti nel sistema in cui è installato SPSS Statistics Server e i file memorizzati nel computer locale che risiedono in cartelle o unità condivise.

Nota: quando si calcolano i punteggi dei dati, il modello verrà applicato alle variabili del file di dati attivo che possiedono lo stesso nome delle variabili presenti nel file del modello. È possibile associare le variabili del modello originale a diverse variabili del file di dati attivo utilizzando la sintassi dei comandi (vedere il comando `MODEL HANDLE`).

Nome. Nome utilizzato per identificare il modello. Le regole relative ai nomi di modello validi sono le stesse che vengono applicate ai nomi di variabile (vedere Nomi delle variabili in Capitolo 5 a pag. 79), con l'aggiunta del carattere $ come primo carattere ammesso). Questo nome potrà essere utilizzato per specificare il modello durante il calcolo del punteggio dei dati con la funzione ApplyModel o StrApplyModel.

File. File XML (PMML) contenente le specifiche del modello.

Missing Values

Questo gruppo di opzioni controlla il trattamento dei valori mancanti, incontrati durante il processo di calcolo del punteggio, per le variabili predittive definite nel modello. Un valore mancante nel contesto del calcolo del punteggio fa riferimento a uno dei seguenti elementi:

- Le variabili predittore non contengono alcun valore. Nel caso delle variabili numeriche, questo indica il valore mancante di sistema e nelle variabili di stringa una stringa nulla.
- Nel modello il valore è stato definito come valore utente non valido per il predittore dato. I valori mancanti definiti dall'utente nell'insieme di dati attivo, ma non nel modello, non vengono considerati come valori mancanti nel calcolo del punteggio.
- La variabile predittore è una variabile categoriale e il valore non corrisponde a una delle categorie definite nel modello.

Usa valore di sostituzione. Tenta di utilizzare la sostituzione del valore durante il calcolo del punteggio nei casi che presentano valori mancanti. Il metodo per determinare quale valore utilizzare per sostituire un valore mancante dipende dal tipo di modello predittivo.

- **SPSS Statistics Modelli.** Per quanto riguarda le variabili indipendenti nei modelli di regressione lineare e discriminanti, se durante la creazione e il salvataggio del modello è stato specificato il valore di sostituzione medio per i valori mancanti, tale valore verrà utilizzato in luogo dei valori mancanti nel calcolo del punteggio, in modo da consentire la continuazione del calcolo stesso. Se il valore medio non è disponibile, verrà restituito il valore mancante di sistema.
- **Modelli AnswerTree & Modelli di comando TREE.** Nei modelli CHAID e CHAID esaustivo, per la variabile di distinzione mancante viene selezionato il nodo figlio più grande, ovvero il nodo figlio con la popolazione più estesa tra i nodi figlio che utilizzano casi di esempio per l'apprendimento. Nei modelli C&RT e QUEST le variabili di distinzione surrogate (se presenti) vengono utilizzate per prime (le suddivisioni surrogate sono suddivisioni che tentano di corrispondere il più possibile alle suddivisioni originali ricorrendo a predittori alternativi). Se non si specifica alcuna suddivisione surrogata o se tutte le variabili di distinzione surrogate risultano mancanti, verrà utilizzato il nodo figlio più grande.
- **Modelli Clementine.** I modelli di regressione lineare vengono gestiti come descritto per i modelli SPSS Statistics. I modelli di regressione logistica vengono gestiti come descritto per i modelli di regressione logistica. I modelli C&RT Tree vengono gestiti come descritto per i modelli C&RT nell'ambito dei modelli AnswerTree.
- **Modelli di regressione logistica.** Per le covariate dei modelli di regressione logistica, se nel modello salvato è stato incluso un valore medio del predittore, tale valore verrà utilizzato in luogo del valore mancante nel calcolo del punteggio, in modo da consentire la continuazione del calcolo stesso. Se il predittore è categoriale (ad esempio un fattore di un modello di regressione logistica) o se il valore medio non è disponibile, verrà restituito il valore mancante di sistema.

Usa mancante di sistema. Restituisce il valore mancante di sistema durante il calcolo del punteggio in un caso che presenta un valore mancante.

Visualizzazione di un elenco di modelli caricati

È possibile ottenere un elenco dei modelli correntemente caricati. Dai menu (disponibili solo in modalità di analisi distribuita) scegliere:

Trasforma
  Prepara modello
    Elenca modelli

Verrà generata una tabella degli handle di modello, contenente un elenco di tutti i modelli correntemente caricati, incluso il nome (definito anche handle di modello) assegnato al modello, il tipo di modello, il percorso del file del modello e il metodo applicato per gestire i valori mancanti.

Figura 8-35
Elenco dei modelli caricati

Model Handles

Name	Type	Input File Name	Validation handling	Missing handling
mregression	NOMREG	C:\models\mregression.xml	ON	SUBSTITUTE
tree	TREE	C:\models\tree.xml	ON	SUBSTITUTE

Funzionalità aggiuntive disponibili con la sintassi dei comandi

Nella finestra di dialogo Carica modello è possibile incollare le proprie selezioni in una finestra di sintassi e modificare la sintassi dei comandi `MODEL HANDLE` risultante. Questo consente di:

- Associare le variabili del modello originale a diverse variabili nel file di dati attivo (utilizzando il sottocomando `MAP`). Per impostazione predefinita, il modello viene applicato alle variabili del file di dati attivo che possiedono lo stesso nome delle variabili presenti nel file del modello.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
9

Gestione e trasformazione di file

I file di dati non sempre sono organizzati nella forma ottimale per esigenze specifiche. Può essere necessario combinare i file di dati, ordinare i dati diversamente, selezionare un sottoinsieme di casi oppure modificare l'unità di analisi raggruppando i dati. In SPSS è disponibile una vasta gamma di funzioni per la trasformazione dei dati, tra cui la possibilità di:

Ordinare i dati. È possibile ordinare i dati in base al valore di una o più variabili.

Trasporre casi e variabili. Con il formato di file di dati di SPSS Statistics, le righe vengono lette come casi e le colonne come variabili. Per i file di dati in cui tale ordine risulta invertito, è possibile scambiare le righe e le colonne e leggere i dati nel formato corretto.

Unire file. È possibile unire due o più file di dati e inoltre combinare i file con le stesse variabili e casi diversi oppure i file con gli stessi casi e variabili diverse.

Selezionare sottoinsiemi di casi. È possibile limitare l'analisi a un sottoinsieme di casi oppure eseguire contemporaneamente più analisi in diversi sottoinsiemi.

Aggregare i dati. È possibile cambiare l'unità di analisi aggregando i dati in base al valore di una o più variabili di raggruppamento.

Ponderare i dati. È possibile pesare i casi per l'analisi in base al valore di una variabile di ponderazione.

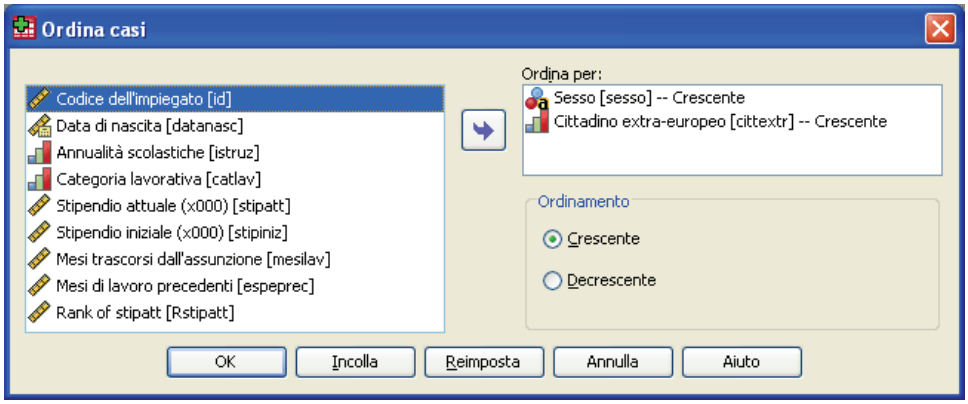
Ristrutturare i dati. È possibile ristrutturare i dati per creare un singolo caso (record) da più casi oppure creare più casi da un singolo caso.

Ordina casi

In questa finestra di dialogo i casi del file di dati vengono ordinati in base ai valori di una o più variabili di ordinamento. È possibile ordinare i casi in ordine crescente o decrescente.

- Se si selezionano più variabili di ordinamento, i casi verranno ordinati in base a ciascuna variabile all'interno delle categorie della variabile precedente nella lista di ordinamento. Se, ad esempio, si seleziona *sesso* come prima variabile di ordinamento e *minoranza* come seconda variabile di ordinamento, i casi verranno classificati per minoranza all'interno di ciascuna categoria di sesso.
- La sequenza di ordinamento è basata sull'ordine definito delle impostazioni internazionali che non è necessariamente uguale all'ordine numerico dei set di caratteri. Le impostazioni internazionali predefinite sono quelle del sistema operativo. Le impostazioni internazionali possono essere controllate tramite l'impostazione Lingua nella scheda Generale della finestra di dialogo Opzioni (menu Modifica).

Figura 9-1
Finestra di dialogo Ordina casi.



Per ordinare i casi

▶ Dai menu, scegliere:

Dati
Ordina casi...

▶ Selezionare una o più variabili di ordinamento.

Ordina variabili

È possibile ordinare le variabili nell'insieme di dati attivo in base ai valori di uno qualsiasi degli attributi delle variabili (ad esempio, nome della variabile, tipo di dati, livello di misurazione), includendo gli attributi di variabile personalizzati.

- I valori possono essere ordinati in ordine crescente o decrescente.
- È possibile salvare l'ordine originale delle variabili (preordinate) in un attributo di variabile personalizzato.
- L'ordinamento in base ai valori degli attributi di variabile personalizzati è limitato agli attributi di variabile personalizzati che risultano visibili nella Visualizzazione variabili.

Per ulteriori informazioni sugli attributi di variabile personalizzati, vedere Attributi variabili personalizzati.

Per ordinare le variabili

Nella Visualizzazione variabili dell'Editor dei dati:

▶ Fare clic con il pulsante destro del mouse sull'intestazione di colonna dell'attributo e dal menu di scelta rapida scegliere Ordinamento crescente oppure Ordinamento decrescente.

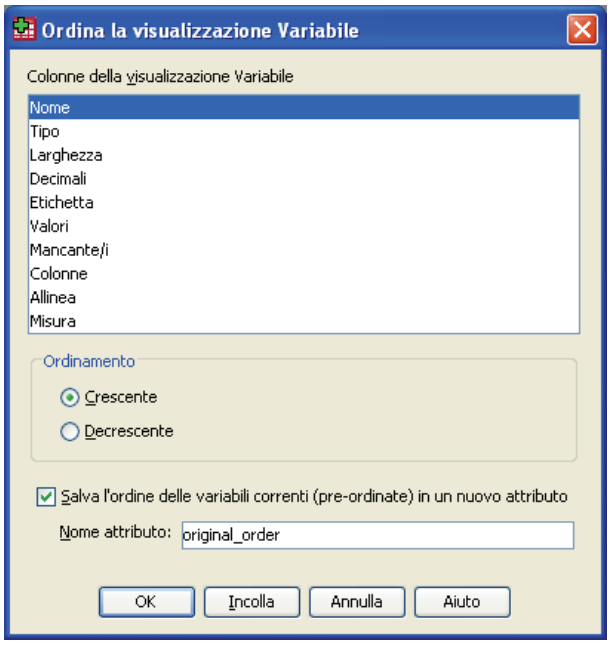
o

▶ In Visualizzazione variabili o Visualizzazione dati, scegliere dai menu:

Dati
Ordina variabili

- Selezionare l'attributo che si desidera utilizzare per ordinare le variabili.
- Selezionare l'ordine (crescente o decrescente).

Figura 9-2
Finestra di dialogo Ordina variabili



- L'elenco degli attributi di variabile corrisponde ai nomi della colonna degli attributi visualizzati nella Visualizzazione variabili dell'Editor dei dati.
- È possibile salvare l'ordine originale delle variabili (preordinate) in un attributo di variabile personalizzato. Per ciascuna variabile, il valore dell'attributo è un valore intero che indica la sua posizione prima dell'ordinamento; pertanto, ordinando le variabili in base al valore di tale attributo personalizzato è possibile ripristinare l'ordine originale della variabile.

Trasponi

L'opzione Trasponi consente di creare un nuovo file di dati in cui le righe e le colonne dei file di dati originali vengono trasposte in modo che i casi (righe) diventino variabili (colonne) e viceversa. Verranno creati automaticamente nuovi nomi di variabile che compariranno in una lista specifica.

- Verrà creata automaticamente una nuova variabile stringa, *caso_lbl*, contenente il nome della variabile originale.
- Se l'insieme di dati attivo contiene una variabile nome o identificativa con valori univoci, è possibile utilizzarla come Variabile nome. I relativi valori verranno utilizzati come nomi di variabili nel file di dati unito. Se si tratta di una variabile numerica, i nomi di variabile iniziano con la lettera *V* seguita dal valore numerico.
- Nel file di dati trasposto i valori mancanti definiti dall'utente verranno convertiti in valori mancanti di sistema. Se si desidera mantenere i valori mancanti come definiti dall'utente, modificare la definizione di valori mancanti in Visualizzazione variabili dell'Editor dei dati.

Per trasporre variabili e casi

▶ Dai menu, scegliere:

Dati
  Trasponi...

▶ Selezionare una o più variabili da trasporre in casi.

Unione di file di dati

È possibile unire i dati contenuti in due file in due diversi modi. È possibile:

- Unire il file di dati attivo con un altro insieme di dati aperto oppure con un file di dati SPSS Statistics che contenga le stesse variabili ma casi diversi.
- Unire il file di dati attivo con un altro insieme di dati aperto oppure con un file di dati SPSS Statistics che contenga gli stessi casi ma variabili diverse.

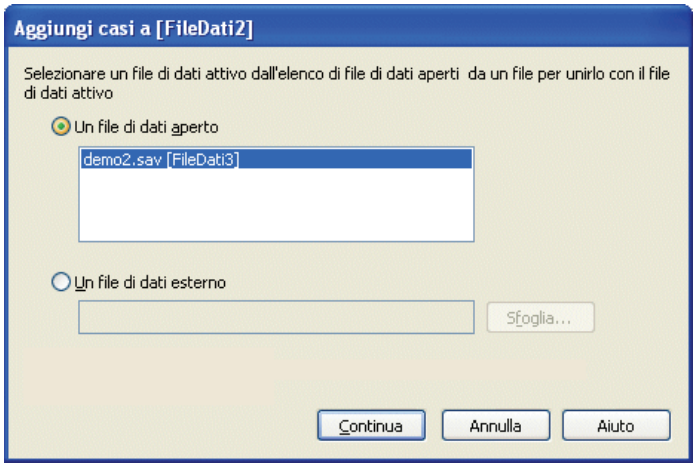
Per unire i file

▶ Dai menu, scegliere:

Dati
  Unisci file

▶ Selezionare Aggiungi casi o Aggiungi variabili.

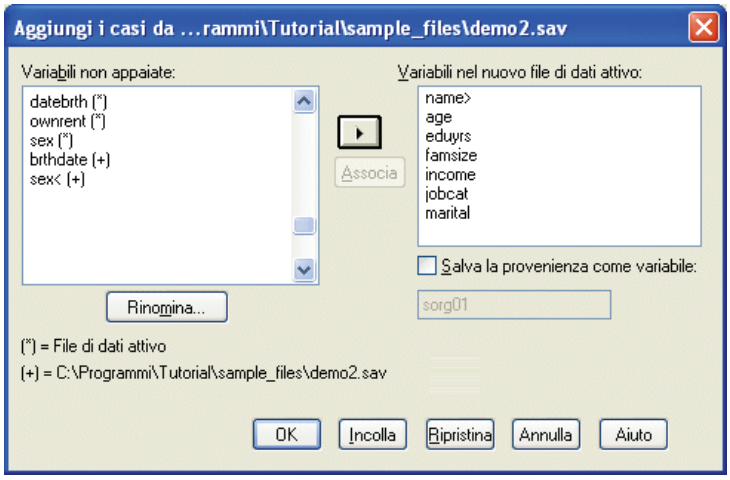
Figura 9-3
Selezione dei file da unire


Aggiungi casi a [FileDati2]
Selezionare un file di dati attivo dall'elenco di file di dati aperti da un file per unirlo con il file di dati attivo
Un file di dati aperto
demo2.sav [FileDati3]
Un file di dati esterno
Sfoglia...
Continua
Annulla
Aiuto

Aggiungi casi

Aggiungi casi unisce il file di dati attivo a un altro insieme di dati o a un file di dati esterno SPSS Statistics che contiene le stesse variabili (colonne) ma casi diversi (righe). È possibile, ad esempio, registrare le stesse informazioni relative ai clienti di due aree di vendita distinte, mantenendo i dati relativi a ciascuna area in un file distinto. Il secondo insieme di dati può essere un file di dati esterno SPSS Statistics o un insieme di dati della sessione corrente.

Figura 9-4
Finestra di dialogo Aggiungi casi



Variabili non appaiate. Le variabili da escludere dal nuovo file di dati unito. Le variabili del file dati attivo sono identificate dall'asterisco (*). Le variabili di altri file di dati sono identificate dal segno più (+). Per impostazione predefinita, la lista contiene:

- Variabili contenute in uno dei due file di dati che non corrispondono a un nome di variabile presente nell'altro file. È possibile creare coppie di variabili non appaiate e quindi includerle nel nuovo file unito.
- Variabili definite come dati numerici in un file e come stringhe nell'altro. Non è possibile unire variabili numeriche con variabili stringa.
- Variabili stringa di lunghezza diversa. La larghezza definita di una variabile stringa deve essere la stessa in entrambi i file di dati.

Variabili in un nuovo file di dati attivo. Le variabili da includere nel nuovo file di dati unito. Per impostazione predefinita, tutte le variabili che corrispondono sia al nome che al tipo di dati (numerico o stringa) vengono incluse nella lista.

- È possibile rimuovere dalla lista le variabili che non si desidera includere nel file unito.
- Le variabili non appaiate incluse nel file unito contengono dati mancanti per i casi del file che non contengono la variabile in questione.

Salva la provenienza come variabile. Indica il file di dati sorgente per ogni caso. valore 0 per i casi provenienti dal file di dati attivo nell'Editor dei dati e valore 1 per i casi provenienti dal file di dati esterno.

Per unire file con le stesse variabili e casi diversi

▶ Aprire almeno uno dei file di dati da unire. Se ci sono più insiemi di dati aperti, impostare uno degli insiemi di dati da unire come insieme di dati attivo. I casi di questo file appariranno per primi nel nuovo file di dati unito.

- Dai menu, scegliere:

  Dati
  Unisci file
  Aggiungi casi...

- Selezionare l'insieme di dati o il file di dati SPSS Statistics esterno da unire all'insieme di dati attivo.

- Rimuovere tutte le variabili indesiderate dall'elenco Variabili nel nuovo file di dati attivo.

- Aggiungere una coppia di variabili inclusa nella lista Variabili non appaiate che rappresenta le stesse informazioni registrate sotto nomi di variabili diverse nei due file. Alla data di nascita, ad esempio, è possibile che in un file corrisponda il nome di variabile *datanasc* e nell'altro *datanasci*.

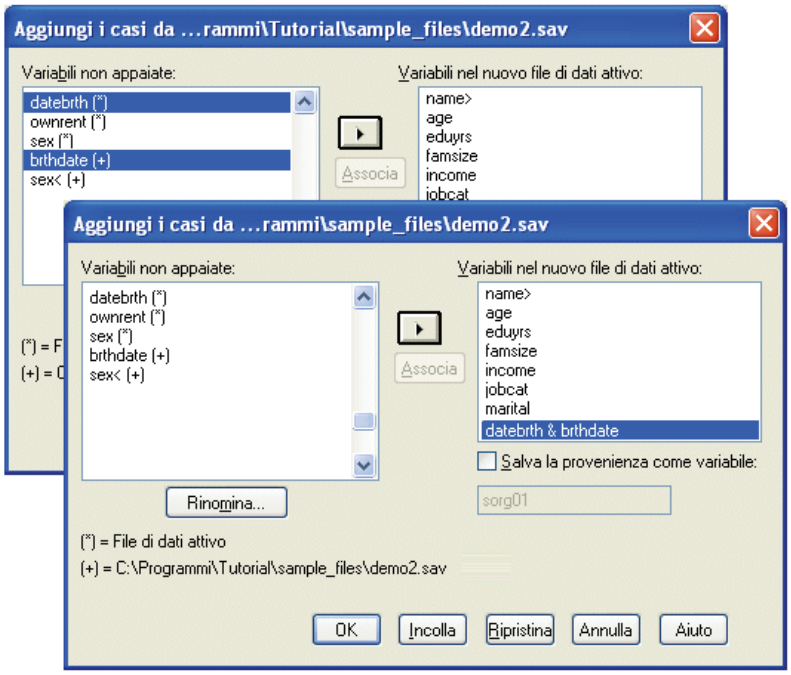
Per selezionare una coppia di variabili non appaiate

- Fare clic su una delle variabili incluse nella lista Variabili non appaiate.

- Fare Ctrl+clic sull'altra variabile nella lista. (Premere Ctrl e contemporaneamente fare clic con il pulsante sinistro del mouse).

- Fare clic su Associa per spostare la coppia di variabili nell'elenco Variabili nel nuovo file dati attivo. Il nome della variabile incluso nel file dati attivo verrà utilizzato come nome di variabile nel file unito.

Figura 9-5
Selezione di coppie di variabili con Ctrl+clic



Aggiungi casi: Rinomina

È possibile rinominare le variabili del file dati attivo o dell'altro file di dati prima di spostarle dalla lista delle variabili non appaiate alla lista delle variabili da includere nel file di dati unito. Rinominando le variabili è possibile:

- Utilizzare un nome di variabile dell'altro file di dati anziché il nome del file dati attivo per le coppie di variabili.
- Includere due variabili con lo stesso nome ma di tipi non appaiati o con stringhe di lunghezza diversa. Ad esempio, per includere sia la variabile numerica *sesso* del file dati attivo che la variabile stringa *sesso* dell'altro file di dati, è innanzitutto necessario rinominare una delle due variabili.

Aggiungi casi: informazioni del dizionario

Le informazioni del dizionario esistenti (etichette delle variabili e dei valori, valori mancanti definiti dall'utente, formati di visualizzazione) nel file dati attivo verranno applicate al file di dati unito.

- Se nel file dati attivo non sono definite informazioni del dizionario per una variabile, verranno utilizzate le informazioni del dizionario del file di dati esterno.
- Se il file dati attivo contiene etichette di valori definite o valori definiti dall'utente per una variabile, verranno ignorate le altre etichette di valori o gli altri valori mancanti definiti dall'utente per tale variabile inclusi nell'altro file di dati.

Unione di più sorgenti dati

La sintassi di comando permette di unire fino a 50 file di dati e/o file. Per ulteriori informazioni, vedere il comando `ADD FILES` in *Command Syntax Reference* (visualizzabile tramite il menu ?).

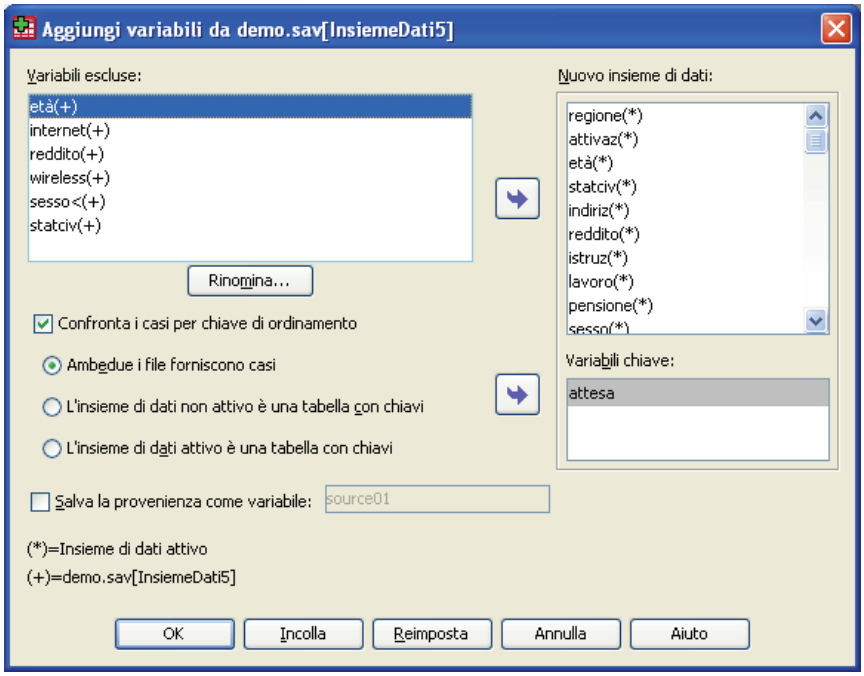
Aggiungi variabili

Aggiungi variabili unisce il file di dati attivo a un altro file di dati aperto o a un file di dati SPSS Statistics esterno che contiene gli stessi casi (righe) ma variabili diverse (colonne). Ad esempio, può essere necessario unire un file di dati contenente i risultati precedenti a un test con un file contenente i risultati successivi al test.

- I casi devono essere ordinati nello stesso modo in entrambi i file di dati.
- Se per confrontare i casi vengono utilizzate una o più variabili chiave, i due file di dati devono essere disposti nell'ordine crescente definito per tali variabili.
- I nomi delle variabili del secondo file di dati uguali ai nomi delle variabili del file dati attivo verranno esclusi per impostazione predefinita in quanto si presume che tali variabili contengano informazioni duplicate.

Salva la provenienza come variabile. Indica il file di dati sorgente per ogni caso. valore 0 per i casi provenienti dal file di dati attivo nell'Editor dei dati e valore 1 per i casi provenienti dal file di dati esterno.

Figura 9-6
Finestra di dialogo Aggiungi variabili



Variabili escluse. Le variabili da escludere dal nuovo file di dati unito. Per impostazione predefinita la lista contiene i nomi delle variabili dell'altro file di dati uguali a quelli del file dati attivo. Le variabili del file dati attivo sono identificate dall'asterisco (*). Le variabili di altri file di dati sono identificate dal segno più (+). Se si desidera includere nel file unito una variabile esclusa con un nome duplicato, è possibile rinominarla e aggiungerla alla lista delle variabili da includere.

Nuovo file di dati attivo. Le variabili da includere nel nuovo file di dati unito. Per impostazione predefinita, tutti i nomi di variabile univoci in entrambi i file di dati verranno inclusi nella lista.

Variabili chiave. Se per alcuni casi inclusi in un file di dati non esistono file corrispondenti nell'altro file di dati (ovvero, se in uno dei file di dati mancano alcuni casi), utilizzare le variabili chiave per identificare e confrontare correttamente i casi di entrambi i file di dati. È inoltre possibile utilizzare le variabili con i file della tabella di consultazione.

- Le variabili chiave devono avere lo stesso nome in entrambi i file di dati.
- Entrambi i file di dati devono essere disposti nell'ordine crescente delle variabili chiave e l'ordine delle variabili della lista Variabili chiave deve corrispondere alla loro sequenza di ordinamento.
- I casi non corrispondenti delle variabili chiave verranno inclusi nel file unito ma non verranno uniti con i casi dell'altro file. I casi non appaiati contengono solo i valori delle variabili incluse nel file da cui provengono; le variabili dell'altro file contengono il valore mancante di sistema.

Il file di dati non attivo o attivo è una tabella con chiave univoca. La chiave univoca, o **tabella di consultazione**, è un file in cui i dati per ciascun "caso" possono essere applicati a più casi inclusi nell'altro file dati. Ad esempio, se un file contiene informazioni sui singoli membri di una famiglia (ad esempio sesso, età, istruzione) e l'altro file contiene informazioni generali sulla famiglia (ad esempio reddito totale, dimensioni, residenza), è possibile utilizzare il file contenente

i dati familiari come tabella di consultazione e applicare i dati comuni sulla famiglia a ciascun membro incluso nel file di dati unito.

Per unire file con gli stessi casi e variabili diverse

- Aprire almeno uno dei file di dati da unire. Se ci sono più insiemi di dati aperti, impostare uno degli insiemi di dati da unire come insieme di dati attivo.
- Dai menu, scegliere:

  Dati
  Unisci file
  Aggiungi variabili...

- Selezionare l'insieme di dati o il file di dati SPSS Statistics esterno da unire all'insieme di dati attivo.

Per selezionare variabili chiave

- Selezionare alcune variabili del file esterno (+) nella lista Variabili escluse.
- Selezionare Confronta i casi per chiave di ordinamento.
- Aggiungere le variabili alla lista Variabili chiave.

  Le variabili chiave devono essere presenti sia nel file di dati attivo che nell'altro file di dati. Entrambi i file di dati devono essere disposti nell'ordine crescente delle variabili chiave e l'ordine delle variabili della lista Variabili chiave deve corrispondere alla loro sequenza di ordinamento.

Aggiungi variabili: Rinomina

È possibile rinominare le variabili del file dati attivo o dell'altro file di dati prima di spostarle dalla lista delle variabili alla lista delle variabili da includere nel file di dati unito. Questo risulta estremamente utile se si desidera includere nei due file due variabili con lo stesso nome che contengono informazioni diverse.

Unione di più sorgenti dati

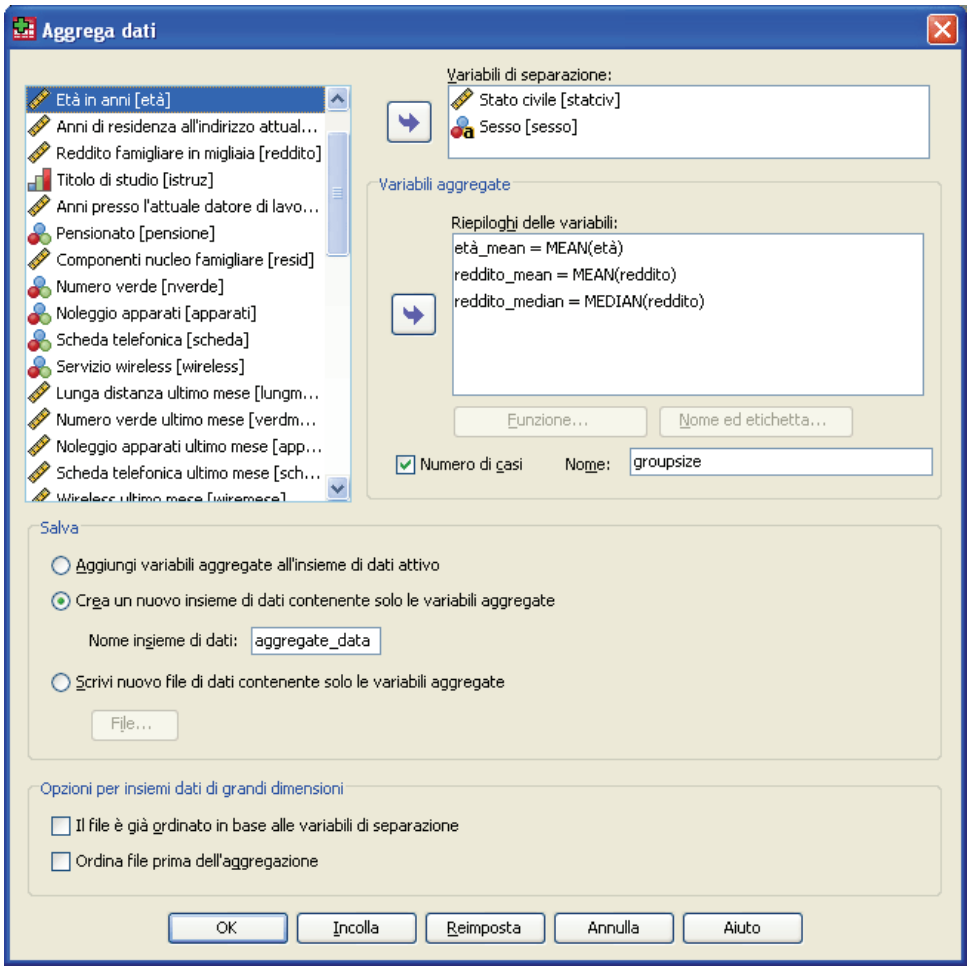
La sintassi di comando permette di unire fino a 50 file di dati e/o file. Per ulteriori informazioni, vedere il comando `MATCH FILES` in *Command Syntax Reference* (visualizzabile tramite il menu ?).

Aggrega dati

Aggrega dati aggrega i gruppi di casi nell'insieme di dati attivo in singoli casi e crea un nuovo file aggregato oppure nuove variabili nell'insieme di dati attivo che contengono i dati aggregati. I casi vengono aggregati in base al valore di *zero* o più variabili di separazione (raggruppamento). Se non vengono specificate variabili di separazione, l'intero insieme di dati è un unico gruppo di separazione.

- Se si crea un nuovo file di dati aggregato, il nuovo file di dati contiene un caso per ciascun gruppo definito per mezzo delle variabili di separazione. Ad esempio, se la variabile di separazione ha due valori, il nuovo file di dati contiene solo due casi. Se non viene specificata una variabile di separazione, il nuovo file di dati conterrà un solo caso.
- Se si aggiungono variabili aggregate all'insieme di dati attivo, questo non viene aggregato. A ciascun caso con gli stessi valori delle variabili di separazione vengono assegnati gli stessi valori delle nuove variabili aggregate. Ad esempio, se il *sesso* è l'unica variabile di separazione, a tutti gli uomini viene assegnato lo stesso valore per la nuova variabile aggregata che rappresenta l'età media. Se non viene specificata una variabile di separazione, tutti i casi ricevono lo stesso valore per una nuova variabile aggregata che rappresenta l'età media.

Figura 9-7
Finestra di dialogo Aggrega dati



Variabili di separazione. I casi vengono raggruppati in base ai valori delle variabili di separazione. Ciascuna combinazione univoca di valori delle variabili di separazione costituisce un gruppo. Quando si crea un nuovo file di dati aggregato, tutte le variabili di separazione vengono salvate nel nuovo file con i nomi e le informazioni del dizionario esistenti. La variabile di separazione, se specificata, può essere di tipo numerico o stringa.

Variabili aggregate. Le variabili sorgente vengono utilizzate con le funzioni aggregate per creare nuove variabili aggregate. Il nome della variabile aggregata è seguito da un'etichetta di variabile facoltativa, dal nome della funzione di aggregazione e dal nome della variabile sorgente tra parentesi.

È possibile annullare i nomi delle variabili aggregate predefinite indicando nuovi nomi di variabile, fornire etichette di variabile descrittive e modificare le funzioni utilizzate per calcolare i valori aggregati. È inoltre possibile creare una variabile che contenga il numero di casi di ciascun gruppo di interruzione.

Per aggregare un file di dati

- Dai menu, scegliere:
  Dati
  Aggrega...
- Se necessario, selezionare le variabili di separazione che definiscono le modalità di raggruppamento dei casi per creare dati aggregati. Se non vengono specificate variabili di separazione, l'intero insieme di dati è un unico gruppo di separazione.
- Selezionare una o più variabili aggregate.
- Selezionare una funzione di aggregazione per ciascuna variabile aggregata.

Salvataggio dei risultati aggregati

È possibile aggiungere variabili aggregate al file di dati attivo oppure creare un nuovo file di dati aggregato.

- **Aggiungi variabili aggregate all'insieme di dati attivo.** All'insieme di dati attivo sono aggiunte nuove variabili basate sulle funzioni di aggregazione. Il file di dati non viene aggregato. A ciascun caso con gli stessi valori delle variabili di separazione vengono assegnati gli stessi valori delle nuove variabili aggregate.
- **Crea un nuovo insieme di dati contenente solo le variabili aggregate.** Salva i dati aggregati in un nuovo insieme di dati nella sessione corrente. Il file di dati include le variabili di separazione che definiscono i casi aggregati e tutte le variabili aggregate definite dalle funzioni aggregate. Questa operazione non ha alcun effetto sull'insieme di dati attivo.
- **Scrivi nuovo file di dati contenente solo le variabili aggregate.** Salva i dati aggregati in un file di dati esterno. Il file include le variabili di separazione che definiscono i casi aggregati e tutte le variabili aggregate definite dalle funzioni aggregate. Questa operazione non ha alcun effetto sull'insieme di dati attivo.

Opzioni di ordinamento per i file di dati di grandi dimensioni

Per i file di grandi dimensioni, può essere più conveniente aggregare i dati pre-ordinati.

Il file è già ordinato in base alle variabili di separazione. Se i dati sono già stati ordinati in base ai valori delle variabili di separazione, l'opzione consente di eseguire la procedura più rapidamente e utilizzare meno memoria. Utilizzare l'opzione con attenzione.

- I dati devono essere ordinati in base ai valori delle variabili di separazione e nello stesso ordine delle variabili di separazione specificate per la procedura Aggrega dati.
- Se si aggiungono variabili al file dei dati attivo, è necessario selezionare questa opzione solo se i dati vengono ordinati in ordine crescente dei valori delle variabili di separazione.

Ordina file prima dell'aggregazione. In casi molto rari con file di dati di grandi dimensioni, potrebbe essere necessario ordinare il file di dati in base ai valori delle variabili di separazione prima di eseguire l'aggregazione. L'opzione è sconsigliata salvo in caso di problemi di prestazioni o relativi alla memoria.

Aggrega dati: Funzione di aggregazione

In questa finestra di dialogo è possibile specificare la funzione da utilizzare per calcolare i valori aggregati per le variabili selezionate nella lista Variabili aggregate nella finestra di dialogo Aggrega dati. Di seguito vengono elencate alcune funzioni di aggregazione:

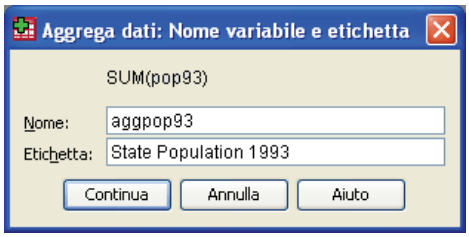
- Funzioni di rappresentazione per variabili numeriche, ad esempio media, mediana, deviazione standard e somma
- Numero di casi, compresi i casi pesati, non pesati, non mancanti e mancanti
- Percentuale o frazioni di valori maggiori o minori del valore specificato
- Percentuale o frazione di valori interni o esterni all'intervallo specificato

Figura 9-8
Finestra di dialogo Funzione di aggregazione

Aggrega dati: Etichetta e nome variabile

La funzione Aggrega dati consente di assegnare nomi di variabili predefiniti alle variabili aggregate del nuovo file di dati. In questa finestra di dialogo è possibile modificare il nome della variabile selezionata nella lista Variabili aggregate e specificare un'etichetta descrittiva per la variabile. Per ulteriori informazioni, vedere Nomi delle variabili in Capitolo 5 a pag. 79.

Figura 9-9
Finestra di dialogo Nome variabile e etichetta

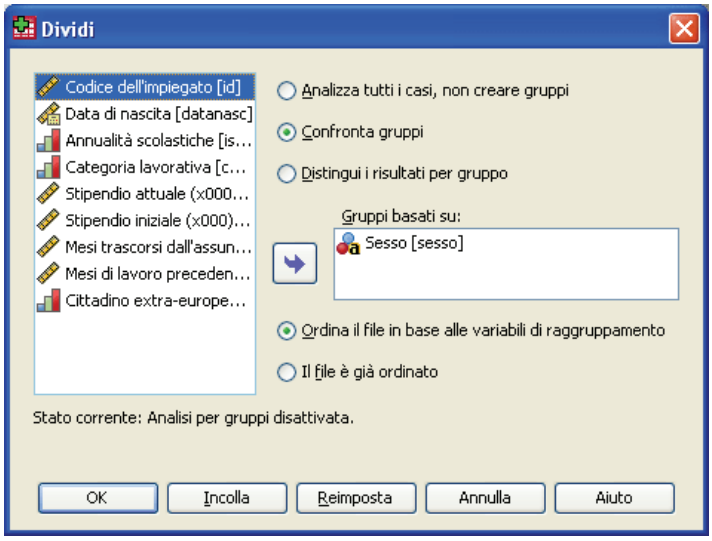


Dividi

L'opzione Dividi consente di suddividere il file di dati in gruppi distinti per l'analisi in base ai valori di una o più variabili di raggruppamento. Se vengono selezionate più variabili di raggruppamento, i casi verranno raggruppati in base a ciascuna variabile all'interno delle categorie della variabile precedente nella lista Gruppi basati su. Se, ad esempio, si seleziona *sesso* come prima variabile di raggruppamento e *minoranza* come seconda variabile di raggruppamento, i casi verranno classificati per minoranza all'interno di ciascuna categoria di sesso.

- È possibile specificare al massimo otto variabili di raggruppamento.
- Ogni gruppo di otto byte di una variabile stringa lunga (variabili stringa composte da più di otto byte) viene contato come una variabile e anche in questo caso vale il limite massimo di otto variabili di raggruppamento.
- I casi devono essere ordinati in base ai valori della variabile di raggruppamento e nello stesso ordine in cui le variabili compaiono nell'elenco Gruppi basati su. Se il file di dati non è già ordinato, selezionare Ordina il file in base alle variabili di raggruppamento.

Figura 9-10
Finestra di dialogo Dividi



Confronta gruppi. Le analisi vengono effettuate sui singoli gruppi, ma i risultati vengono riportati in una sola tabella riepilogativa per facilitare i confronti. Per quanto riguarda le tabelle pivot, viene creata una sola tabella e ciascuna variabile di distinzione può essere spostata tra le dimensioni della tabella. Per quanto riguarda i grafici, viene creato un grafico distinto per ciascun gruppo di analisi; i grafici verranno visualizzati insieme nel Viewer.

Distingui i risultati per gruppo. Tutti i risultati di ciascuna procedura verranno visualizzati separatamente per ciascun gruppo di analisi.

Per suddividere un file di dati in blocchi di dati per l'analisi

▶ Dai menu, scegliere:

Dati
  Dividi...

▶ Selezionare Confronta gruppi o Distingui i risultati per gruppo.

▶ Selezionare una o più variabili di raggruppamento.

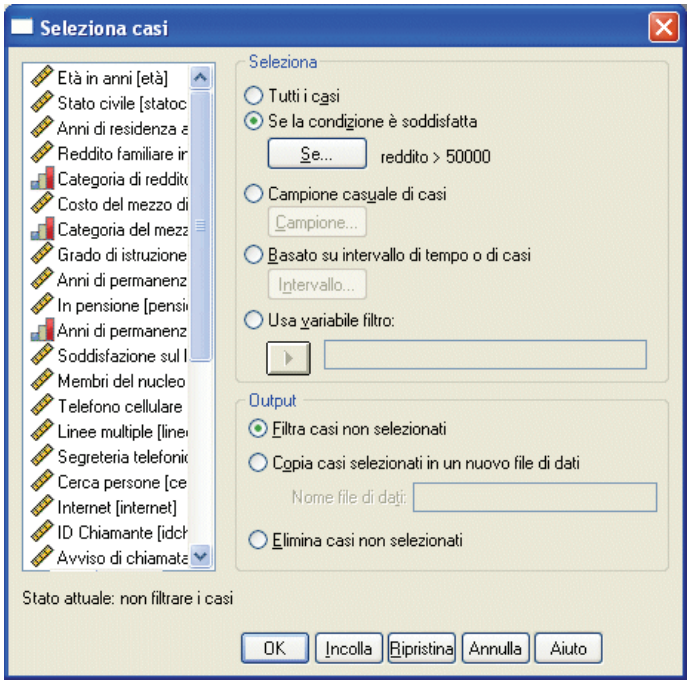
Seleziona casi

L'opzione Seleziona casi fornisce diversi metodi di selezione di un sottoinsieme di casi in base a criteri che includono variabili ed espressioni complesse. È inoltre possibile selezionare un campione casuale di casi. I criteri in base ai quali viene definito un sottogruppo includono:

- Valori di variabili e intervalli
- Intervalli di date e orari
- Numeri di casi (righe)
- Espressioni aritmetiche

- Espressioni logiche
- Funzioni

Figura 9-11
Finestra di dialogo Seleziona casi



Tutti i casi. Sopprime il filtro corrente e usa tutti i casi nelle analisi successive.

Se la condizione è soddisfatta. Usa un'espressione condizionale per selezionare i casi. Se il risultato dell'espressione logica è vero, il caso verrà selezionato. I casi che non soddisfano la condizione impostata vengono esclusi dal campione.

Campione casuale di casi. Seleziona un campione casuale basato su una percentuale approssimativa o un numero esatto di casi.

Basato su intervallo di tempo o di casi. Seleziona i casi in base ad un intervallo di casi o di date/ore.

Usa variabile filtro. Usa la variabile numerica selezionata dal file di dati come variabile di filtro. Vengono esclusi dal campione i casi con valore 0 o mancante per questa variabile.

Output

Questa sezione controlla il trattamento dei casi non selezionati. Per il trattamento dei casi non selezionati sono disponibili le opzioni seguenti:

- **Rimuovi casi non selezionati.** I casi non selezionati non saranno inclusi nell'analisi, ma rimarranno nell'insieme di dati. e sarà possibile utilizzarli successivamente nella sessione se si disattivano le funzioni di filtro. Se si seleziona un campione casuale o si selezionano casi in base a un'espressione logica, verrà creata una variabile denominata *filtro_$* con un valore 1 per i casi selezionati e un valore 0 per i casi non selezionati.

- **Copia casi selezionati in un nuovo insieme di dati.** I casi selezionati vengono copiati in un nuovo insieme di dati senza modificare quello originale. I casi non selezionati non vengono inclusi nel nuovo insieme di dati e vengono lasciati nello stato originale nell'insieme di dati originale.
- **Elimina casi non selezionati.** I casi non selezionati verranno cancellati dall'insieme di dati. Per recuperare i casi cancellati è necessario uscire dal file senza salvare le modifiche e quindi riaprire il file. Se si salvano le modifiche al file di dati, l'eliminazione dei casi sarà definitiva.

Nota: Se si eliminano i casi non selezionati e si salva il file, non sarà possibile recuperare i casi.

Per selezionare un sottoinsieme di casi

▶ Dai menu, scegliere:
Dati
Seleziona casi...

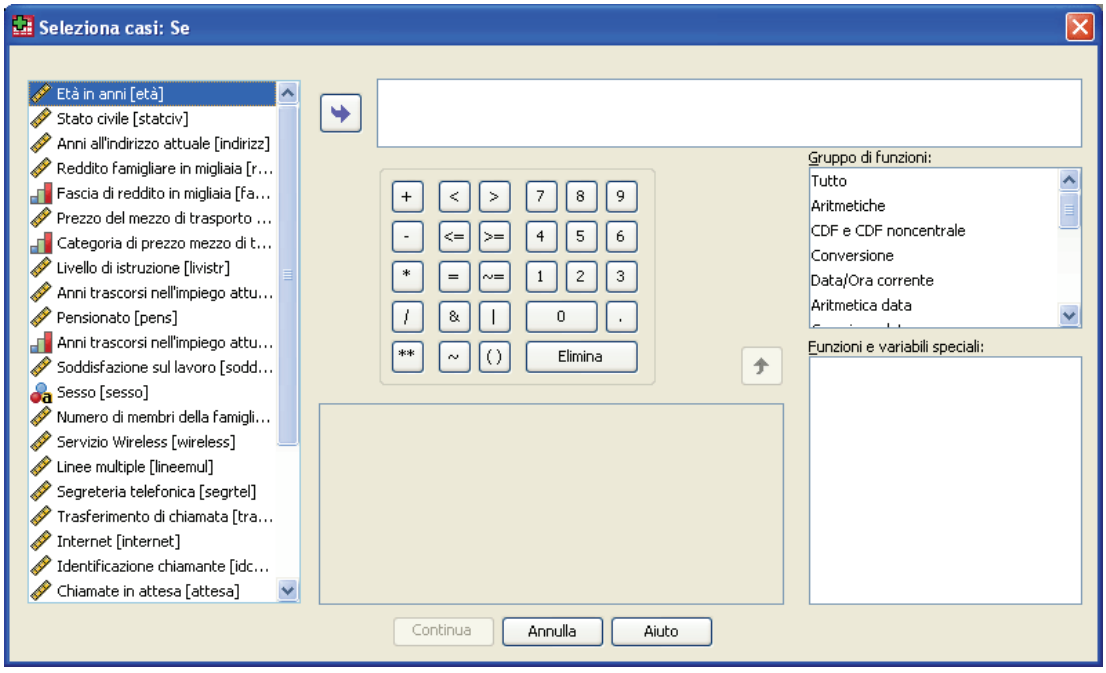
▶ Selezionare uno dei metodi di selezione dei casi.

▶ Specificare i criteri di selezione dei casi.

Seleziona casi: Se

In questa finestra di dialogo è possibile selezionare sottoinsiemi di casi utilizzando espressioni logiche. Un'espressione logica restituisce un valore *vero*, *falso*, o *mancante* per ciascun caso.

Figura 9-12
Finestra di dialogo Seleziona casi: Se



- Se il risultato di un'espressione logica è *vero*, il caso verrà incluso nel sottoinsieme selezionato.
- Se il risultato di un'espressione logica è *falso* o *mancante*, il caso non viene incluso nel sottoinsieme selezionato.

- La maggior parte delle espressioni condizionali utilizza uno o più dei sei operatori relazionali (<, >, <=, >=, = e ~=) della calcolatrice.
- Le espressioni logiche possono includere nomi di variabili, costanti, operatori aritmetici, funzioni numeriche e di altro tipo, variabili logiche e operatori relazionali.

Seleziona casi: Campione casuale

In questa finestra di dialogo è possibile selezionare un campione casuale in base a una percentuale approssimativa o a un numero esatto di casi. Il campionamento viene eseguito senza sostituzione: in tal modo è impossibile selezionare lo stesso caso più di una volta.

Figura 9-13
Finestra di dialogo Seleziona casi: Campione casuale

Approssimativamente. Genera un campione casuale che include approssimativamente la percentuale di casi specificata. Poiché per ciascun caso viene eseguito un processo indipendente di decisione pseudo-casuale, l'equivalenza tra la percentuale di casi selezionati e la percentuale specificata può essere solo approssimativa. Maggiore è il numero di casi inclusi nel data file e maggiore sarà l'approssimazione della percentuale di casi selezionati rispetto alla percentuale specificata.

Esattamente. Un numero di casi specificato dall'utente. È necessario specificare anche il numero di casi in base al quale generare il campione. Il secondo numero deve essere uguale o minore del numero totale di casi del file di dati. Se questo numero supera il numero totale di casi del file di dati, in proporzione il campione conterrà un numero di dati inferiore a quello specificato.

Seleziona casi: Intervallo

In questa finestra di dialogo è possibile selezionare casi in base a un intervallo di numeri di casi o a un intervallo di date e orari.

- Gli intervalli di casi si basano sul numero di riga visualizzato nell'Editor dei dati.
- Gli intervalli di date e orari sono disponibili solo per i **dati di serie storiche** con variabili della data definite (menu Dati, Definisci date).

Figura 9-14
Finestra di dialogo Seleziona casi: intervallo per intervalli di casi (nessuna variabile di data definita)

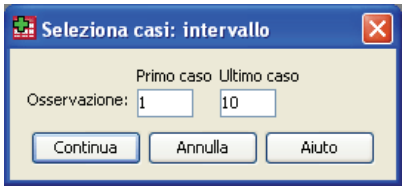
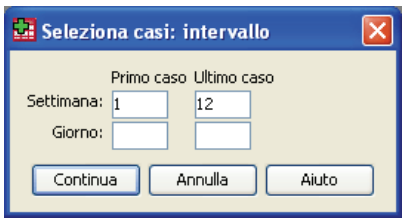


Figura 9-15
Finestra di dialogo Seleziona casi: intervallo per dati di serie storiche con variabili di data definite

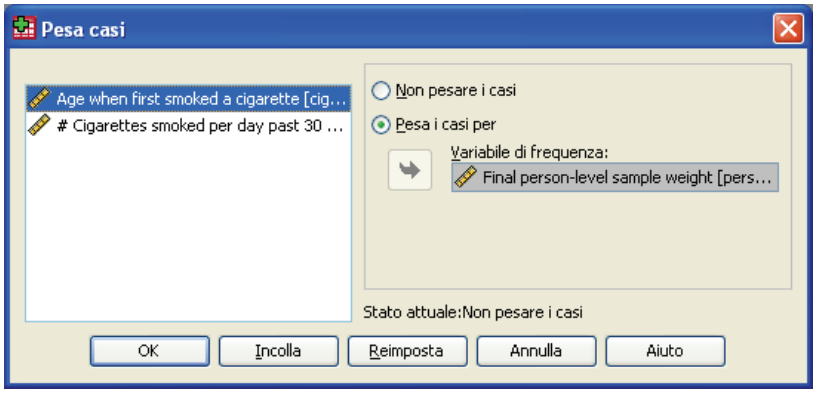


Pesa casi

L'opzione Pesa casi consente di assegnare ai casi pesi diversi (mediante replicazione simulata) per l'analisi statistica.

- I valori della variabile di ponderazione devono indicare il numero di osservazioni rappresentate dai singoli casi del file di dati.
- I casi con valori uguali a zero, negativi o mancanti per la variabile di ponderazione verranno esclusi dall'analisi.
- I valori frazionari sono validi e alcune procedure, quali Frequenze, Tavole di contingenza e Tabelle personalizzate, utilizzeranno i valori di ponderazione frazionari. Tuttavia la maggior parte delle procedure tratta la variabile di ponderazione come un peso replicazione e arrotonderà semplicemente i pesi frazionari all'intero più vicino. Alcune procedure ignorano completamente la variabile di ponderazione e questo limite viene evidenziato nella documentazione delle specifiche procedure.

Figura 9-16
Finestra di dialogo Pesa casi



Una volta applicata, una variabile di ponderazione rimane valida finché non ne viene selezionata un'altra o si disattiva la ponderazione. Quando si salva un file di dati ponderato, le informazioni di ponderazione verranno salvate insieme al file. È possibile disattivare la ponderazione in qualsiasi momento, anche dopo aver salvato il file in forma ponderata.

Ponderazione nelle tavole di contingenza. Nella procedura Tavole di contingenza sono disponibili diverse opzioni per la gestione dei pesi dei casi. Per ulteriori informazioni, vedere Tavole di contingenza: Visualizzazione cella in Capitolo 18 a pag. 308.

Ponderazione in grafici a dispersione e istogrammi. Per i grafici a dispersione e gli istogrammi è disponibile un'opzione per attivare e disattivare i pesi dei casi. Tale opzione, tuttavia, non produce alcun effetto sui casi con valore negativo, uguale a 0 o mancante per la variabile di ponderazione. Tali casi verranno esclusi dal grafico anche se la ponderazione viene disattivata dal suo interno.

Per pesare i casi

▶ Dai menu, scegliere:

Dati
  Pesa casi...

▶ Selezionare Pesa i casi per.

▶ Selezionare una variabile di frequenza.

I valori della variabile di frequenza vengono utilizzati come pesi dei casi. Ad esempio un caso con un valore 3 per la variabile di frequenza rappresenterà 3 casi nel file di dati ponderato.

Ristrutturazione di dati

La Ristrutturazione di dati guidata consente di ristrutturare i dati per la procedura che si desidera utilizzare. La procedura guidata sostituisce il file corrente con un nuovo file ristrutturato e offre le funzioni seguenti:

- Ristruttura le variabili selezionate in casi
- Ristruttura i casi selezionati in variabili
- Trasponi tutti i dati

Per ristrutturare i dati

▶ Dai menu, scegliere:

Dati
  Ristruttura...

▶ Selezionare il tipo di ristrutturazione che si desidera eseguire.

▶ Selezionare i dati da ristrutturare.

Se lo si desidera, è possibile:

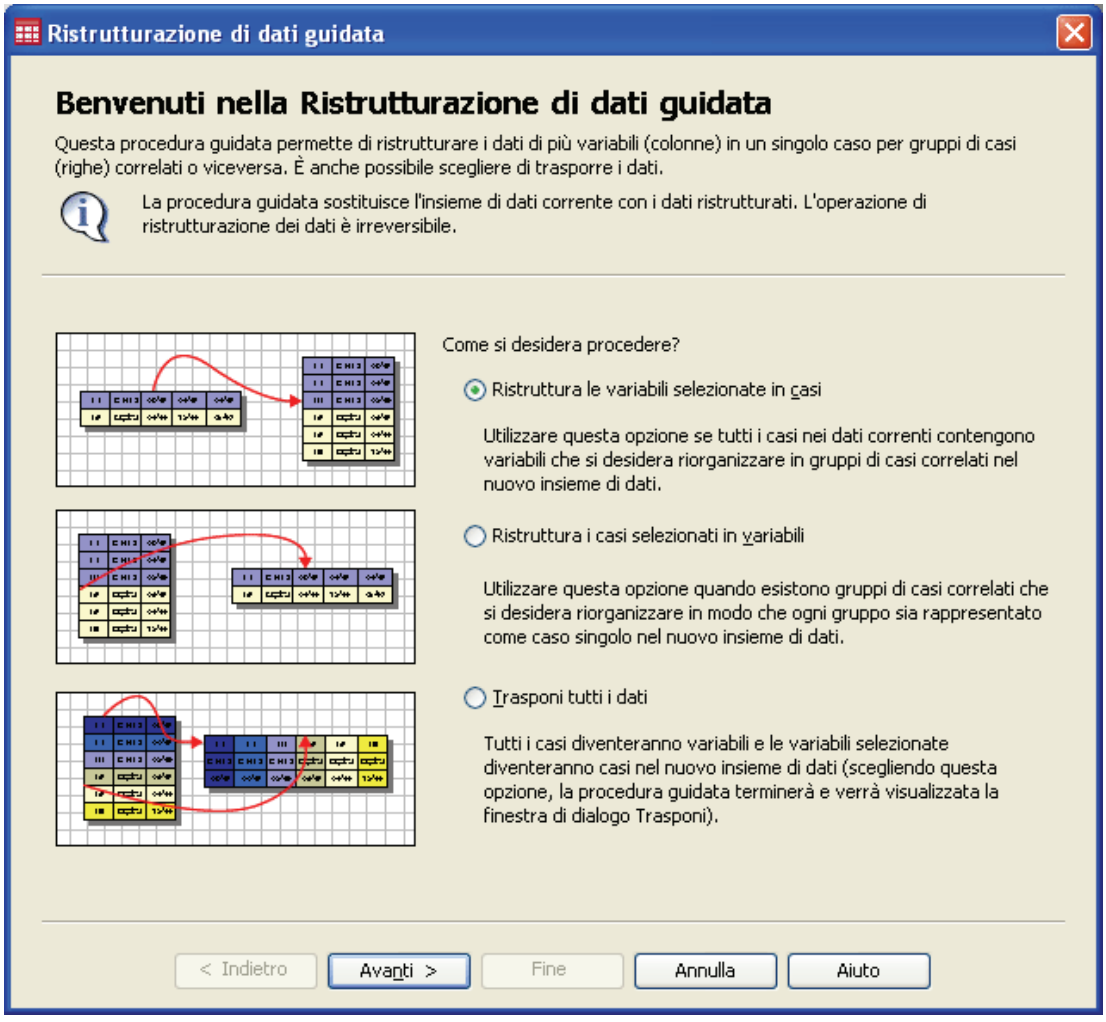
- Creare variabili di identificazione, che consentono di risalire da un valore nel nuovo file a un valore nel file originale

- Ordinare i dati prima della ristrutturazione
- Definire le opzioni per il nuovo file
- Incollare la sintassi dei comandi in una finestra di sintassi

Ristrutturazione di dati guidata: Seleziona tipo

La Ristrutturazione di dati guidata consente di ristrutturare i dati in uso. Nella prima finestra di dialogo selezionare il tipo di ristrutturazione che si desidera eseguire.

Figura 9-17
Ristrutturazione di dati guidata


Ristrutturazione di dati guidata
Benvenuti nella Ristrutturazione di dati guidata
Questa procedura guidata permette di ristrutturare i dati di più variabili (colonne) in un singolo caso per gruppi di casi (righe) correlati o viceversa. È anche possibile scegliere di trasporre i dati.
La procedura guidata sostituisce l'insieme di dati corrente con i dati ristrutturati. L'operazione di ristrutturazione dei dati è irreversibile.
Come si desidera procedere?
Ristruttura le variabili selezionate in casi
Utilizzare questa opzione se tutti i casi nei dati correnti contengono variabili che si desidera riorganizzare in gruppi di casi correlati nel nuovo insieme di dati.
Ristruttura i casi selezionati in variabili
Utilizzare questa opzione quando esistono gruppi di casi correlati che si desidera riorganizzare in modo che ogni gruppo sia rappresentato come caso singolo nel nuovo insieme di dati.
Trasponi tutti i dati
Tutti i casi diventeranno variabili e le variabili selezionate diventeranno casi nel nuovo insieme di dati (scegliendo questa opzione, la procedura guidata terminerà e verrà visualizzata la finestra di dialogo Trasponi).
< Indietro
Avanti >
Fine
Annulla
Aiuto

- **Ristruttura le variabili selezionate in casi.** Selezionare questa opzione se nei dati sono presenti gruppi di colonne correlate che si desidera visualizzare come gruppi di righe nel nuovo file di dati. Se si sceglie questa opzione, verranno visualizzati i passi per **Variabili in casi**.

- **Ristruttura i casi selezionati in variabili.** Selezionare questa opzione se nei dati sono presenti gruppi di righe correlate che si desidera visualizzare come gruppi di colonne nel nuovo file di dati. Se si sceglie questa opzione, verranno visualizzati i passi per **Casi in variabili**.
- **Trasponi tutti i dati.** Selezionare questa opzione se si desidera trasporre i dati. Tutte le righe diventeranno colonne e tutte le colonne diventeranno righe nei nuovi dati. Se si sceglie questa opzione, la Ristrutturazione di dati guidata verrà chiusa e verrà aperta la finestra di dialogo Trasponi dati.

Scelta della modalità di ristrutturazione dei dati

Una **variabile** contiene le informazioni che si desidera analizzare, ad esempio una misura o un punteggio. Un **caso** è un'osservazione, ad esempio un individuo. In una struttura di dati *semplice* ogni variabile è una singola colonna di dati e ogni caso è una singola riga. Se, ad esempio, si misurano i punteggi dei test relativi a tutti gli studenti di una classe, tutti i valori vengono visualizzati in una singola colonna e per ogni studente è presente una singola riga.

L'analisi dei dati comporta in genere l'analisi della modalità di variazione di una variabile in base a una condizione specifica, che può essere rappresentata da una cura sperimentale specifica, da informazioni demografiche, da un punto temporale o da altro. Nell'analisi dei dati, le condizioni prese in considerazione vengono in genere denominate **fattori**. Se si analizzano i fattori, la struttura dei dati sarà *complessa*. Le informazioni relative a una variabile possono essere presenti in più colonne dei dati, ad esempio in una colonna per ogni livello di un fattore, oppure le informazioni relative a un caso possono essere presenti in più righe, ad esempio una riga per ogni livello di un fattore. La Ristrutturazione di dati guidata semplifica la ristrutturazione dei file con una struttura di dati complessa.

Le scelte eseguite nella procedura guidata variano in base alla struttura del file corrente e alla struttura che si desidera creare nel nuovo file.

Disposizione dei dati nel file corrente. I dati correnti possono essere disposti in modo tale che i fattori vengano registrati in una variabile *distinta* (in gruppi di casi) o *con* la variabile (in gruppi di variabili).

- **Gruppi di casi.** Si supponga che le variabili e le condizioni del file corrente siano registrate in colonne distinte. Ad esempio:

var	fattore
8	1
9	1
3	2
1	2

In questo esempio le prime due righe sono un **gruppo di casi** perché sono correlate e contengono i dati relativi allo stesso livello di fattore. Nell'analisi dei dati SPSS Statistics, il fattore viene in genere definito **variabile di raggruppamento** quando i dati sono strutturati in questo modo.

- **Gruppi di colonne.** Si supponga che le variabili e le condizioni del file corrente siano registrate nella stessa colonna. Ad esempio:

var_1	var_2
8	3
9	1

In questo esempio, le due colonne sono un **gruppo di variabili** perché sono correlate. Contengono i dati della stessa variabile, *var_1* per il livello di fattore 1 e *var_2* per il livello di fattore 2. Nell'analisi dei dati SPSS Statistics, il fattore viene in genere definito **misura ripetuta** quando i dati sono strutturati in questo modo.

Disposizione dei dati nel nuovo file. È in genere determinata dalla procedura che si desidera utilizzare per l'analisi dei dati.

- **Procedure che richiedono gruppi di casi.** È necessario che i dati siano strutturati in gruppi di casi se si desidera eseguire analisi che richiedono una variabile di raggruppamento, ad esempio *univariata*, *multivariata* e *componenti della varianza* con la procedura Modello lineare generalizzato, Modelli misti, Cubi OLAP e *campioni indipendenti* con le procedure Test T o Test non parametrici. Se i dati correnti sono strutturati in gruppi di variabili e si desidera eseguire le analisi, selezionare Ristruttura le variabili selezionate in casi.
- **Procedure che richiedono gruppi di variabili.** È necessario che i dati siano strutturati in gruppi di variabili se si desidera analizzare le misure ripetute, ad esempio *misure ripetute* con la procedura Modello lineare generalizzato, l'analisi *covariata dipendente dal tempo* con l'analisi Regressione di Cox, *campioni appaiati* con Test T o *campioni dipendenti* con Test non parametrici. Se i dati correnti sono strutturati in gruppi di casi e si desidera eseguire le analisi, selezionare Ristruttura i casi selezionati in variabili.

Esempio di Variabili in casi

In questo esempio i punteggi dei test vengono registrati in colonne distinte per ogni fattore, *A* e *B*.

Figura 9-18
Dati correnti per Variabili in casi

	punteg_a	punteg_b
1	1014.00	864.00
2	684.00	636.00
3	810.00	638.00

Si desidera eseguire un test *t* per campioni indipendenti ed è disponibile un gruppo di colonne che include *punteggio_a* e *punteggio_b*, ma non è disponibile la **variabile di raggruppamento** richiesta dalla procedura. Selezionare Ristruttura le variabili selezionate in casi nella Ristrutturazione di dati guidata, ristrutturare un gruppo di variabili in una nuova variabile denominata *punteggio* e creare un indice denominato *gruppo*. Il nuovo file di dati è illustrato nella figura seguente.

Figura 9-19
Nuovi dati ristrutturati per Variabili in casi

	grouppo	punteggio
1	PUNTEG_A	1014.00
2	PUNTEG_B	864.00
3	PUNTEG_A	684.00
4	PUNTEG_B	636.00
5	PUNTEG_A	810.00
6	PUNTEG_B	638.00

Quando si esegue il test *t* per campioni indipendenti sarà possibile utilizzare *gruppo* come variabile di raggruppamento.

Esempio di Casi in variabili

In questo esempio i punteggi dei test vengono registrati due volte per ogni soggetto, —prima e dopo il trattamento.

Figura 9-20
Dati correnti per Casi in variabili

	id	punteg	tempo
1	1	1014.00	prec
2	1	864.00	succ
3	2	684.00	prec
4	2	636.00	succ

Si desidera eseguire un test *t* per campioni appaiati. I dati sono strutturati in gruppi di casi, ma non sono disponibili le **misure ripetute** per le coppie di variabili richieste dalla procedura. Selezionare Ristruttura i casi selezionati in variabili nella Ristrutturazione di dati guidata, utilizzare *id* per identificare i gruppi di righe nei dati correnti e *orario* per creare il gruppo di variabili nel nuovo file.

Figura 9-21
Nuovi dati ristrutturati per Casi in variabili

	id	succ	prec
1	1	864.00	1014.00
2	2	636.00	684.00

Nella procedura Test *t* per campioni appaiati, è possibile utilizzare *prima* e *dopo* come coppia di variabili.

Ristrutturazione di dati guidata (Variabili in casi): Numero di gruppi di variabili

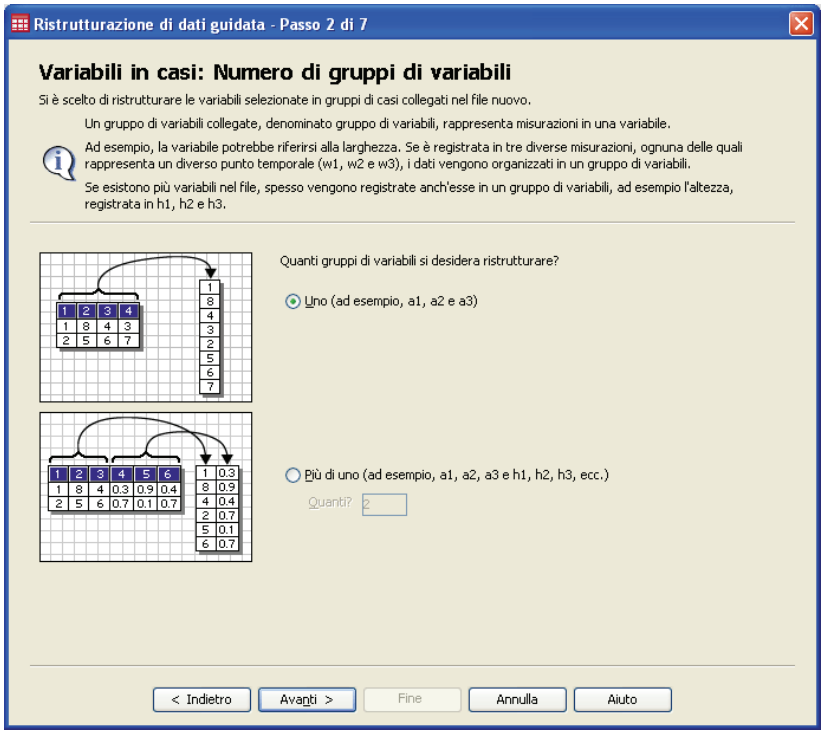
Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe e consente di scegliere il numero di gruppi di variabili del file corrente che si desidera ristrutturare nel nuovo file.

Numero dei gruppi di variabili nel file corrente. Stabilire quanti gruppi di variabili sono presenti nei dati correnti. Un gruppo di colonne correlate, denominato **gruppo di variabili**, registra le misure ripetute della stessa variabile in colonne distinte. Ad esempio, se nei dati correnti sono presenti le tre colonne *a1*, *a2* e *a3* che registrano **ampiezza**, significa che è presente un gruppo di

variabili. Se sono presenti le tre colonne aggiuntive *h1*, *h2* e *h3* che registrano **altezza**, significa che sono disponibili due gruppi di variabili.

Numero dei gruppi di variabili nel nuovo file. Stabilire il numero dei gruppi di variabili che si desidera rappresentare nel nuovo file di dati, tenendo presente che non è necessario ristrutturare tutti i gruppi di variabili nel nuovo file.

Figura 9-22
Ristrutturazione di dati guidata: Numero di gruppi di variabili, Passo 2

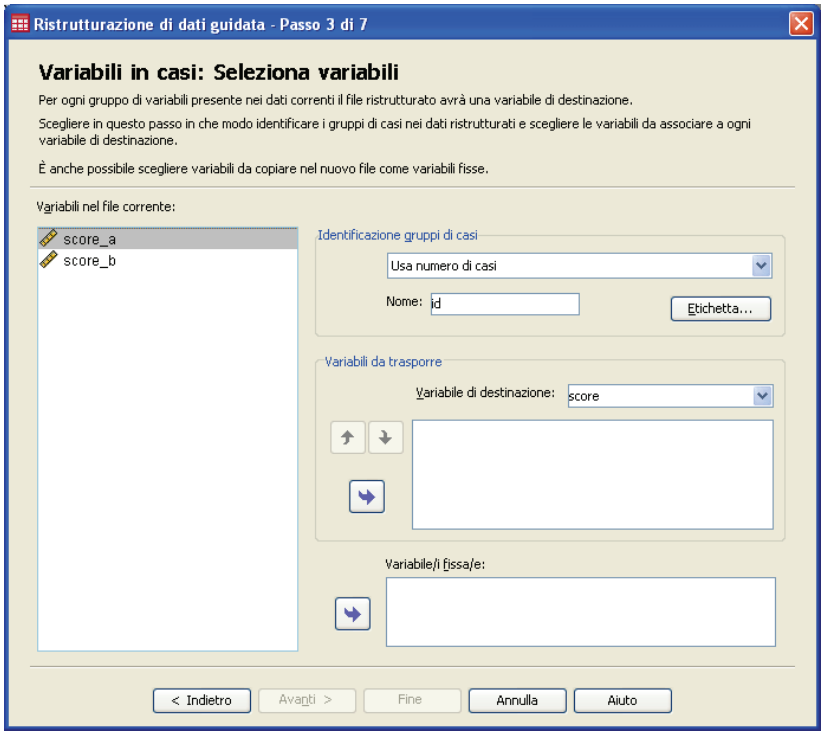

Ristrutturazione di dati guidata - Passo 2 di 7
Variabili in casi: Numero di gruppi di variabili
Si è scelto di ristrutturare le variabili selezionate in gruppi di casi collegati nel file nuovo.
Un gruppo di variabili collegate, denominato gruppo di variabili, rappresenta misurazioni in una variabile.
Ad esempio, la variabile potrebbe riferirsi alla larghezza. Se è registrata in tre diverse misurazioni, ognuna delle quali rappresenta un diverso punto temporale (w1, w2 e w3), i dati vengono organizzati in un gruppo di variabili.
Se esistono più variabili nel file, spesso vengono registrate anch'esse in un gruppo di variabili, ad esempio l'altezza, registrata in h1, h2 e h3.
Quanti gruppi di variabili si desidera ristrutturare?
Uno (ad esempio, a1, a2 e a3)
Più di uno (ad esempio, a1, a2, a3 e h1, h2, h3, ecc.)
Quanti?
< Indietro
Avanti >
Fine
Annulla
Aiuto

- **Uno.** Nel nuovo file verrà creata una singola variabile ristrutturata in base a un gruppo di variabili del file corrente.
- **Più di uno.** Nel nuovo file verranno create più variabili ristrutturate. Il numero specificato influisce sul passo successivo, nel quale viene creato automaticamente il numero specificato di nuove variabili.

Ristrutturazione di dati guidata (Variabili in casi): Seleziona Variabili

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe e fornisce informazioni su come utilizzare le variabili del file corrente nel nuovo file. È inoltre possibile creare una variabile che identifica le righe nel nuovo file.

Figura 9-23
Ristrutturazione di dati guidata: Selezione delle variabili, Passo 3


Ristrutturazione di dati guidata - Passo 3 di 7
Variabili in casi: Seleziona variabili
Per ogni gruppo di variabili presente nei dati correnti il file ristrutturato avrà una variabile di destinazione.
Scegliere in questo passo in che modo identificare i gruppi di casi nei dati ristrutturati e scegliere le variabili da associare a ogni variabile di destinazione.
È anche possibile scegliere variabili da copiare nel nuovo file come variabili fisse.
Variabili nel file corrente:
score_a
score_b
Identificazione gruppi di casi
Usa numero di casi
Nome: id
Etichetta...
Variabili da trasporre
Variabile di destinazione: score
Variabile/i fissa/e:
< Indietro
Avanti >
Fine
Annulla
Aiuto

Modalità di identificazione delle nuove righe. È possibile creare nel nuovo file di dati una variabile che identifica la riga del file corrente utilizzata per creare un gruppo di nuove righe. L'identificatore può essere rappresentato da un numero sequenziale di caso oppure dai valori della variabile. Per definire la variabile di identificazione nel nuovo file, utilizzare i controlli disponibili in Identificazione gruppi di casi. Fare clic su una cella per modificare il nome predefinito della variabile e specificare un'etichetta descrittiva per la variabile di identificazione.

Elementi da ristrutturare nel nuovo file. Nel passo precedente è stato specificato il numero dei gruppi di variabili che si desidera ristrutturare. Per ogni gruppo è stata creata una nuova variabile e i valori per il gruppo di variabili verranno visualizzati in tale variabile all'interno del nuovo file. Per definire la variabile ristrutturata nel nuovo file, utilizzare i controlli disponibili in Variabili da trasporre.

Per specificare una variabile ristrutturata

- Inserire le variabili che costituiscono il gruppo di variabili che si desidera trasformare nell'elenco Variabili da trasporre. Tutte le variabili nel gruppo devono essere dello stesso tipo (numerico o di stringa).

  È possibile includere la stessa variabile più di una volta nel gruppo di variabili (le variabili vengono copiate anzichè spostate dall'elenco sorgente delle variabili) e i corrispondenti valori verranno ripetuti nel nuovo file.

Per specificare più variabili ristrutturate

- Selezionare la prima variabile di destinazione che si desidera definire dall'elenco a discesa.

- Inserire le variabili che costituiscono il gruppo di variabili che si desidera trasformare nell'elenco Variabili da trasporre. Tutte le variabili nel gruppo devono essere dello stesso tipo (numerico o di stringa). Nel gruppo di variabili è possibile includere la stessa variabile più volte. Una variabile viene copiata invece che spostata dall'elenco di variabili di origine e i relativi valori vengono ripetuti nel nuovo file.

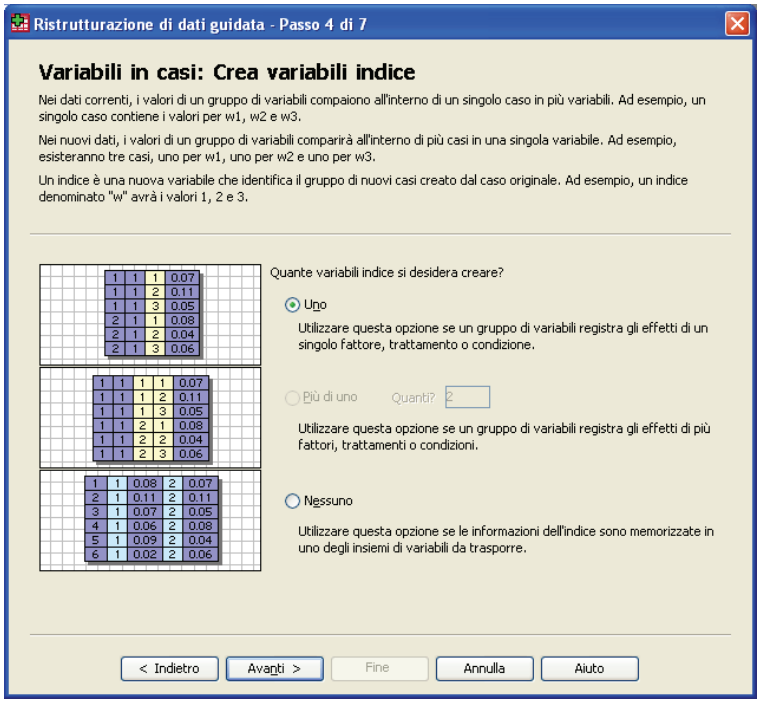
- Selezionare la successiva variabile di destinazione che si desidera definire e ripetere il processo di selezione della variabile per tutte le variabili di destinazione disponibili.
  - Benché sia possibile includere la stessa variabile più di una volta nello stesso gruppo di variabili di destinazione, non è possibile includere la stessa variabile in più di un gruppo di variabili di destinazione.
  - Ogni elenco relativo al gruppo di variabili di destinazione deve contenere lo stesso numero di variabili. Le variabili elencate più di una volta vengono incluse nel conteggio.
  - Il numero di gruppi di variabili di destinazione è determinato dal numero di gruppi di variabili specificati nel passo precedente. Qui è possibile modificare i nomi predefiniti delle variabili, ma è necessario ritornare al passo precedente per modificare il numero dei gruppi di variabili da ristrutturare.
  - È necessario definire i gruppi di variabili (selezionando le variabili dall'elenco sorgente) per tutte le variabili di destinazione disponibili prima di procedere al passo successivo.

Elementi da copiare nel nuovo file. È possibile copiare nel nuovo file le variabili che non vengono ristrutturate, i cui valori verranno propagati nelle nuove righe. Spostare le variabili che si desidera copiare nel nuovo file nell'elenco Variabile/i fissa/e.

Ristrutturazione di dati guidata (Variabili in casi): Crea variabili Variabili

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe e consente di scegliere se creare variabili indice. Un indice è una nuova variabile che identifica in modo sequenziale un gruppo di righe in base alla variabile originale per cui è stata creata la nuova riga.

Figura 9-24
Ristrutturazione di dati guidata: Creazione delle variabili indice, Passo 4


Ristrutturazione di dati guidata - Passo 4 di 7
Variabili in casi: Crea variabili indice
Nei dati correnti, i valori di un gruppo di variabili compaiono all'interno di un singolo caso in più variabili. Ad esempio, un singolo caso contiene i valori per w1, w2 e w3.
Nei nuovi dati, i valori di un gruppo di variabili comparirà all'interno di più casi in una singola variabile. Ad esempio, esisteranno tre casi, uno per w1, uno per w2 e uno per w3.
Un indice è una nuova variabile che identifica il gruppo di nuovi casi creato dal caso originale. Ad esempio, un indice denominato "w" avrà i valori 1, 2 e 3.
Quante variabili indice si desidera creare?
Uno
Utilizzare questa opzione se un gruppo di variabili registra gli effetti di un singolo fattore, trattamento o condizione.
Più di uno
Quanti?
Utilizzare questa opzione se un gruppo di variabili registra gli effetti di più fattori, trattamenti o condizioni.
Nessuno
Utilizzare questa opzione se le informazioni dell'indice sono memorizzate in uno degli insiemi di variabili da trasporre.
< Indietro
Avanti >
Fine
Annulla
Aiuto

Numero delle variabili indice nel nuovo file. Le variabili indice possono essere utilizzate come variabili di raggruppamento nelle procedure. Nella maggior parte dei casi è sufficiente una singola variabile indice. Se tuttavia i gruppi di variabili nel file corrente rappresentano più livelli di fattore, può essere opportuno prevedere più indici.

- **Uno.** Verrà creata una singola variabile indice.
- **Più di uno.** Verranno creati più indici e sarà possibile specificare il numero degli indici che si desidera creare. Il numero specificato influisce sul passo successivo, nel quale viene creato automaticamente il numero di indici specificato.
- **Nessuno.** Selezionare questa opzione se non si desidera creare variabili indice nel nuovo file.

Esempio di indice singolo per Variabili in casi

Nei dati correnti sono presenti il gruppo di variabili *ampiezza* e il fattore *orario*. L'ampiezza è stata misurata tre volte e registrata in *a1*, *a2* e *a3*.

Figura 9-25
Dati correnti per un Indice

	soggetto	a1	a2	a3
1	1	6.70	4.30	5.70
2	2	7.10	5.90	5.60

Il gruppo di variabili verrà ristrutturato in una singola variabile, *ampiezza*, e verrà creato un singolo indice numerico. I nuovi dati sono illustrati nella tabella seguente.

Figura 9-26
Nuovi dati ristrutturati con un indice

	soggetto	indice	ampiezza
1	1	1	6.70
2	1	2	4.30
3	1	3	5.70
4	2	1	7.10
5	2	2	5.90
6	2	3	5.60

Indice inizia con 1 e viene incrementato per ogni variabile presente nel gruppo. Si riavvia ogni volta che trova una nuova riga nel file originale. Ora è possibile usare *indice* nelle procedure che richiedono una variabile di raggruppamento.

Esempio di due indici per Variabili in casi

Se un gruppo di variabili registra più fattori, è possibile creare più indici. I dati correnti devono tuttavia essere disposti in modo tale che i livelli del primo fattore rappresentino un indice primario all'interno del quale i livelli dei fattori successivi eseguono dei cicli. Nei dati correnti è presente il gruppo di variabili *ampiezza* e i due fattori *A* e *B*. I dati sono disposti in modo tale che i livelli del fattore *B* eseguano cicli all'interno dei livelli del fattore *A*.

Figura 9-27
Dati correnti per due indici

	soggetto	a_a1b1	a_a1b2	a_a2b1	a_a2b2
1	1	5.50	6.40	5.80	5.90
2	2	7.40	7.10	5.60	6.70

Il gruppo di variabili verrà ristrutturato in una singola variabile, *ampiezza*, e verranno creati due indici. I nuovi dati sono illustrati nella tabella seguente.

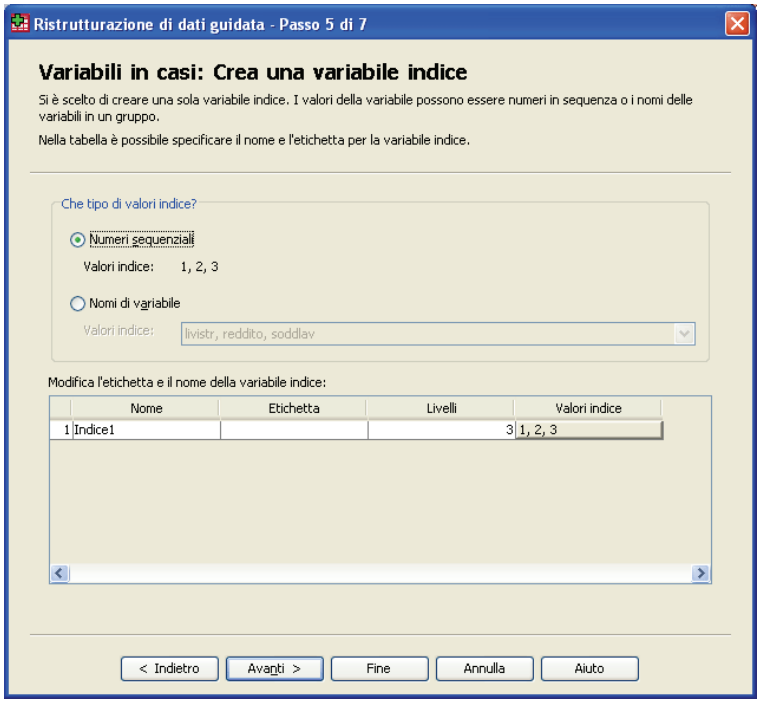
Figura 9-28
Nuovi dati ristrutturati con due indici

	soggetto	indice_a	indice_b	ampiezza
1	1	1	1	5.50
2	1	1	2	6.40
3	1	2	1	5.80
4	1	2	2	5.90
5	2	1	1	7.40
6	2	1	2	7.10
7	2	2	1	5.60
8	2	2	2	6.70

Ristrutturazione di dati guidata (Variabili in casi): Crea una variabile indice

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe e di creare una variabile indice e consente di stabilire i valori desiderati per la variabile indice. I valori possono essere rappresentati da numeri sequenziali o dai nomi delle variabili presenti in un gruppo di variabili originale. È inoltre possibile specificare un nome e un'etichetta per la nuova variabile indice.

Figura 9-29
Ristrutturazione di dati guidata: Creazione di una variabile indice, Passo 5


Ristrutturazione di dati guidata - Passo 5 di 7
Variabili in casi: Crea una variabile indice
Si è scelto di creare una sola variabile indice. I valori della variabile possono essere numeri in sequenza o i nomi delle variabili in un gruppo.
Nella tabella è possibile specificare il nome e l'etichetta per la variabile indice.
Che tipo di valori indice?
Numeri sequenziali
Valori indice: 1, 2, 3
Nomi di variabile
Valori indice: livistr, reddito, soddlav
Modifica l'etichetta e il nome della variabile indice:
Nome
Etichetta
Livelli
Valori indice
1 Indice1 3 1, 2, 3
< Indietro
Avanti >
Fine
Annulla
Aiuto

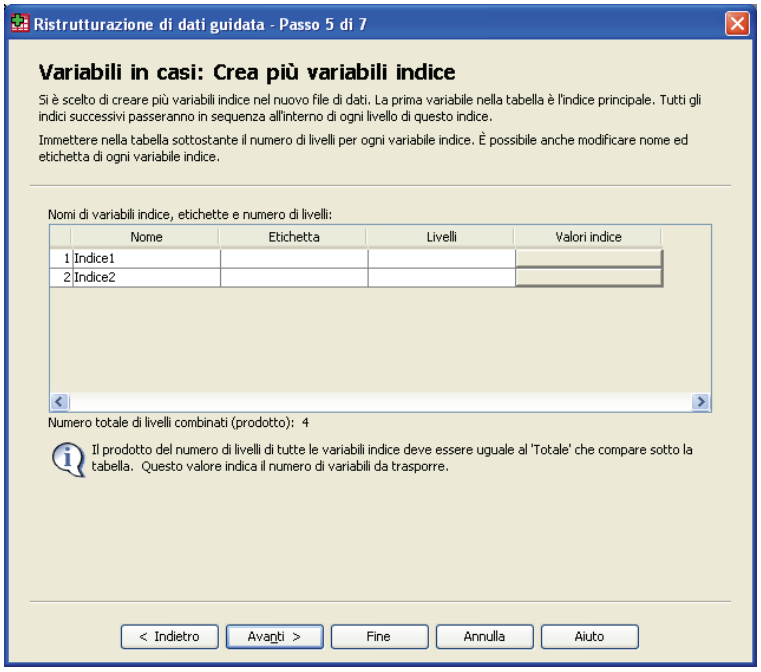
Per ulteriori informazioni, vedere Esempio di indice singolo per Variabili in casi a pag. 206.

- **Numeri sequenziali.** Come valori dell'indice verranno assegnati automaticamente numeri sequenziali.
- **Nomi delle variabili.** Come valori dell'indice verranno utilizzati i nomi del gruppo di variabili selezionato. Selezionare un gruppo di variabili dall'elenco.
- **Nomi ed etichette.** Fare clic su una cella per modificare il nome predefinito della variabile e specificare un'etichetta descrittiva per la variabile indice.

Ristrutturazione di dati guidata (Variabili in casi): Crea più variabili indice

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe e di creare più variabili indice e consente di specificare il numero di livelli per ogni variabile indice. È inoltre possibile specificare un nome e un'etichetta per la nuova variabile indice.

Figura 9-30
Ristrutturazione di dati guidata: Creazione di più variabili indice, Passo 5


Ristrutturazione di dati guidata - Passo 5 di 7
Variabili in casi: Crea più variabili indice
Si è scelto di creare più variabili indice nel nuovo file di dati. La prima variabile nella tabella è l'indice principale. Tutti gli indici successivi passeranno in sequenza all'interno di ogni livello di questo indice.
Immettere nella tabella sottostante il numero di livelli per ogni variabile indice. È possibile anche modificare nome ed etichetta di ogni variabile indice.
Nomi di variabili indice, etichette e numero di livelli:
Nome
Etichetta
Livelli
Valori indice
1 Indice1
2 Indice2
Numero totale di livelli combinati (prodotto): 4
Il prodotto del numero di livelli di tutte le variabili indice deve essere uguale al 'Totale' che compare sotto la tabella. Questo valore indica il numero di variabili da trasporre.
< Indietro
Avanti >
Fine
Annulla
Aiuto

Per ulteriori informazioni, vedere Esempio di due indici per Variabili in casi a pag. 207.

Numero di livelli registrati nel file corrente. Prendere in considerazione il numero dei livelli di fattore registrati nei dati correnti. Un **livello** definisce un gruppo di casi per i quali sono valide le stesse condizioni. Se sono presenti più fattori, i dati correnti devono essere disposti in modo tale che i livelli del primo fattore rappresentino un indice primario all'interno del quale i livelli dei fattori successivi eseguono cicli.

Numero dei livelli nel nuovo file. Specificare il numero dei livelli per ogni indice. I valori per più variabili indice sono sempre rappresentati da numeri sequenziali. I valori iniziano da 1 e vengono incrementati per ogni livello. Il primo indice incrementa il più lento e l'ultimo incrementa il più veloce.

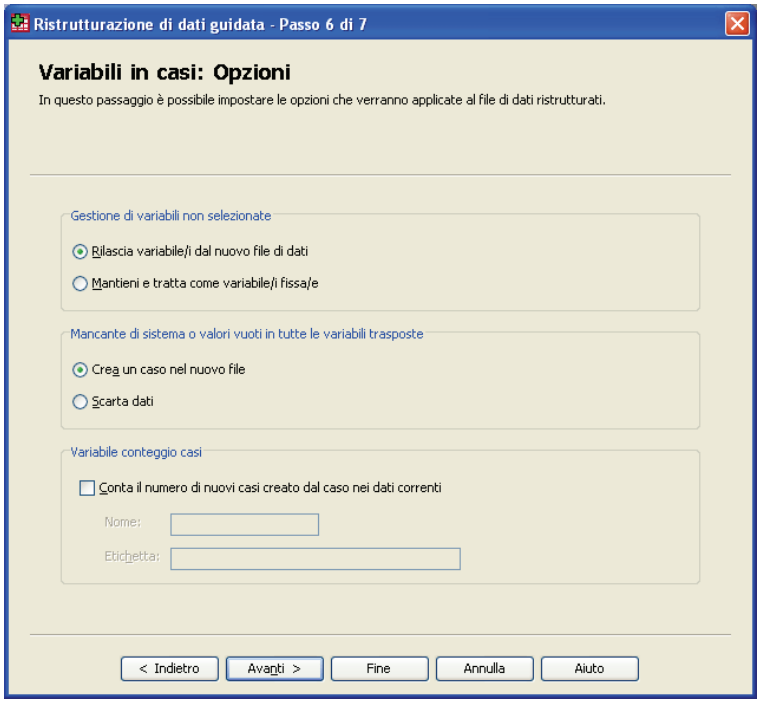
Numero totale di livelli combinati. Non è possibile creare più livelli di quelli presenti nei dati correnti. Poiché i dati ristrutturati conterranno una riga per ogni combinazione di trattamenti, la procedura guidata verifica il numero dei livelli creati e confronta il prodotto dei livelli creati con il numero delle variabili presenti nei gruppi di variabili, che devono corrispondere.

Nomi ed etichette. Fare clic su una cella per modificare il nome predefinito della variabile e specificare un'etichetta descrittiva per le variabili indice.

Ristrutturazione di dati guidata (Variabili in casi): opzioni

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe e consente di specificare le opzioni per il nuovo file ristrutturato.

Figura 9-31
Ristrutturazione di dati guidata: Opzioni, Passo 6


Ristrutturazione di dati guidata - Passo 6 di 7
Variabili in casi: Opzioni
In questo passaggio è possibile impostare le opzioni che verranno applicate al file di dati ristrutturati.
Gestione di variabili non selezionate
Rilascia variabile/i dal nuovo file di dati
Mantieni e tratta come variabile/i fissa/e
Mancante di sistema o valori vuoti in tutte le variabili trasposte
Crea un caso nel nuovo file
Scarta dati
Variabile conteggio casi
Conta il numero di nuovi casi creato dal caso nei dati correnti
Nome:
Etichetta:
< Indietro
Avanti >
Fine
Annulla
Aiuto

Esclusione delle variabili non selezionate. Nel passo 3 relativo alla selezione delle variabili sono stati selezionati dai dati correnti i gruppi di variabili da ristrutturare, le variabili da copiare e una variabile di identificazione. I dati delle variabili selezionate verranno visualizzati nel nuovo file. Se nei dati correnti sono presenti altre variabili, è possibile scegliere di escluderle oppure di mantenerle.

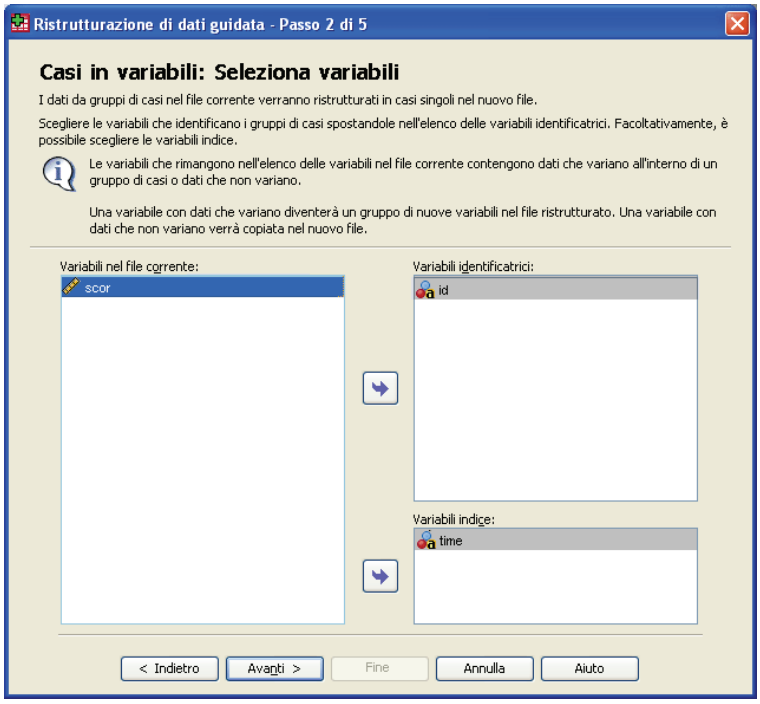
Conservazione dei dati mancanti. La procedura guidata verifica se in ogni nuova riga potenziale sono presenti **valori nulli** ovvero valori mancanti di sistema o vuoti. Si può scegliere di mantenere o escludere le righe che contengono valori nulli.

Creazione di una variabile conteggio. È possibile creare nel nuovo file una **variabile conteggio** che contiene il numero delle nuove righe create da una riga nei dati correnti. Le variabili conteggio possono risultare utili se si sceglie di escludere i valori nulli dal nuovo file perché in questo modo è possibile creare un numero diverso di nuove righe per una riga specifica dei dati correnti. Fare clic su una cella per modificare il nome predefinito della variabile e specificare un'etichetta descrittiva per la variabile conteggio.

Ristrutturazione di dati guidata (Casi in variabili): Seleziona Variabili

Nota: Questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di casi in colonne e fornisce informazioni su come utilizzare le variabili del file corrente nel nuovo file.

Figura 9-32
Ristrutturazione di dati guidata: Selezione delle variabili, Passo 2


Ristrutturazione di dati guidata - Passo 2 di 5
Casi in variabili: Seleziona variabili
I dati da gruppi di casi nel file corrente verranno ristrutturati in casi singoli nel nuovo file.
Scegliere le variabili che identificano i gruppi di casi spostandole nell'elenco delle variabili identificatrici. Facoltativamente, è possibile scegliere le variabili indice.
Le variabili che rimangono nell'elenco delle variabili nel file corrente contengono dati che variano all'interno di un gruppo di casi o dati che non variano.
Una variabile con dati che variano diventerà un gruppo di nuove variabili nel file ristrutturato. Una variabile con dati che non variano verrà copiata nel nuovo file.
Variabili nel file corrente:
scor
Variabili identificatrici:
id
Variabili indice:
time
< Indietro
Avanti >
Fine
Annulla
Aiuto

Identificazione dei gruppi di casi nei dati correnti. Un **gruppo di casi** è un gruppo di righe correlate dal fatto che misurano la stessa unità di osservazione, ad esempio un individuo o un'istituzione. È necessario specificare le variabili del file corrente che identificano i gruppi di casi, in modo tale che sia possibile consolidare ogni gruppo in una singola riga del nuovo file. Spostare le variabili che identificano i gruppi di casi nell'elenco Variabile/i identificatrice/i. Per identificare i gruppi di casi vengono utilizzate le variabili per la suddivisione del file di dati corrente. Ogni volta che viene rilevata una nuova combinazione di valori di identificazione verrà creata una nuova riga e pertanto i casi del file corrente dovrebbero essere ordinati in base ai valori delle variabili di identificazione e allo stesso ordine delle variabili nell'elenco Variabile/i identificatrice/i. Se il file di dati corrente non è già ordinato, è possibile eseguirne l'ordinamento nel passo successivo.

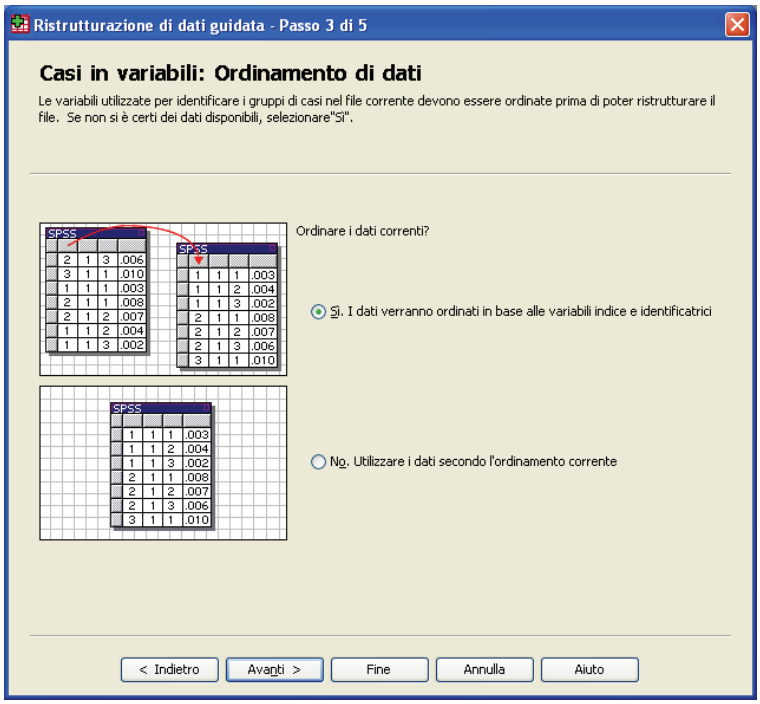
Creazione dei nuovi gruppi di variabili nel nuovo file. Nei dati originali, una variabile viene visualizzata in una singola colonna, mentre nel nuovo file di dati, la stessa variabile viene visualizzata in più colonne nuove. Le **variabili indice** sono le variabili dei dati correnti che vengono utilizzate per creare le nuove colonne. I dati ristrutturati conterranno una nuova variabile per ogni valore univoco di tali colonne. Spostare le variabili che verranno utilizzate per creare i nuovi gruppi di variabili nell'elenco Variabile/i indice. Quando vengono visualizzate le opzioni, è inoltre possibile scegliere di ordinare le nuove colonne in base all'indice.

Cosa accade alle altre colonne. La procedura guidata stabilisce automaticamente cosa accade alle altre variabili rimaste nell'elenco relativo al file corrente. Controlla ogni variabile per verificare se i valori dei dati variano all'interno di un gruppo di casi. In caso affermativo, ristruttura i valori in un gruppo di variabili del nuovo file. In caso contrario, copia i valori nel nuovo file.

Ristrutturazione di dati guidata (Casi in variabili): Ordinamento Dati

Nota: Questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di casi in colonne e consente di stabilire se si desidera ordinare il file corrente prima di ristrutturarlo. Ogni volta che viene rilevata una nuova combinazione di valori di identificazione, verrà creata una nuova riga ed è pertanto importante che i dati vengano ordinati in base alle variabili che identificano i gruppi di casi.

Figura 9-33
Ristrutturazione di dati guidata: Ordinamento dei dati, Passo 3


Ristrutturazione di dati guidata - Passo 3 di 5
Casi in variabili: Ordinamento di dati
Le variabili utilizzate per identificare i gruppi di casi nel file corrente devono essere ordinate prima di poter ristrutturare il file. Se non si è certi dei dati disponibili, selezionare"Sì".
Ordinare i dati correnti?
Sì. I dati verranno ordinati in base alle variabili indice e identificatrici
No. Utilizzare i dati secondo l'ordinamento corrente
< Indietro
Avanti >
Fine
Annulla
Aiuto

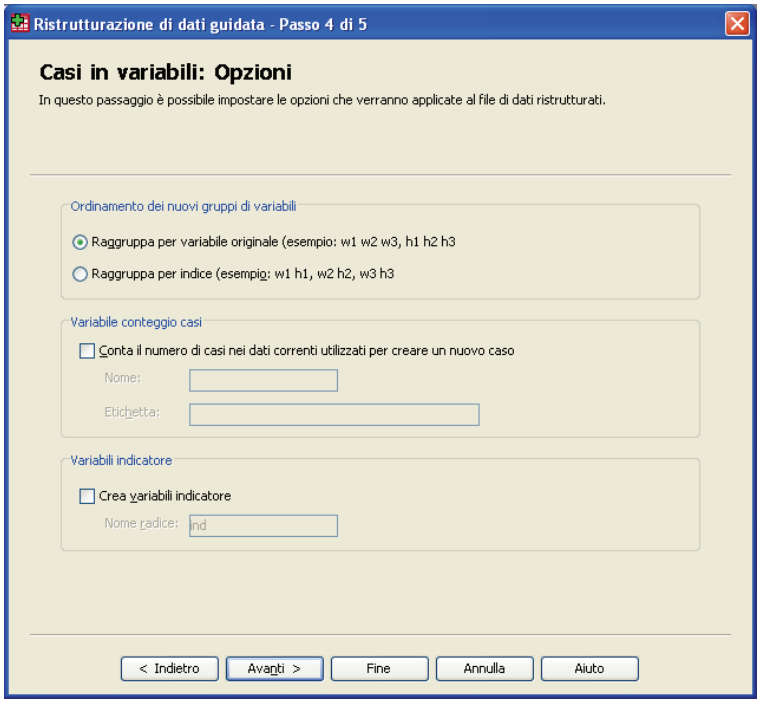
Ordinamento delle righe nel file corrente. Prendere in considerazione la modalità di ordinamento dei dati correnti e le variabili utilizzate per identificare i gruppi di casi, specificate nel passo precedente.

- **Sì.** I dati correnti verranno automaticamente ordinati in base alle variabili di identificazione, nello stesso ordine in cui le variabili vengono visualizzate nell'elenco Variabile/i identificatrice/i del passo precedente. Selezionare questa opzione se i dati non sono ordinati in base alle variabili di identificazione o in caso di incertezze. Questa opzione rende necessario un passaggio distinto dei dati, ma garantisce che le righe vengano ordinate correttamente per la ristrutturazione.
- **No.** I dati correnti non vengono ordinati. Selezionare questa opzione se si è certi che i dati correnti siano ordinati in base alle variabili che identificano i gruppi di casi.

Ristrutturazione di dati guidata (Casi in variabili): opzioni

Nota: Questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di casi in colonne e consente di specificare le opzioni per il nuovo file ristrutturato.

Figura 9-34
Ristrutturazione di dati guidata: Opzioni, Passo 4


Ristrutturazione di dati guidata - Passo 4 di 5
Casi in variabili: Opzioni
In questo passaggio è possibile impostare le opzioni che verranno applicate al file di dati ristrutturati.
Ordinamento dei nuovi gruppi di variabili
Raggruppa per variabile originale (esempio: w1 w2 w3, h1 h2 h3
Raggruppa per indice (esempio: w1 h1, w2 h2, w3 h3
Variabile conteggio casi
Conta il numero di casi nei dati correnti utilizzati per creare un nuovo caso
Nome:
Etichetta:
Variabili indicatore
Crea variabili indicatore
Nome radice: ind
< Indietro
Avanti >
Fine
Annulla
Aiuto

Ordinamento dei nuovi gruppi di variabili nel nuovo file

- **Per variabile.** Le nuove variabili create da una variabile originale vengono raggruppate insieme.
- **Tramite indice.** Le variabili vengono raggruppate in base ai valori delle variabili indice.

Esempio. Le variabili da ristrutturare sono *a* e *h* e l'indice è *mese*:

a	h	mese

Il raggruppamento per variabile restituisce i risultati seguenti:

a.gen	a.feb	h.gen

Il raggruppamento tramite indice restituisce i risultati seguenti:

a.gen	h.gen	a.feb

Creazione di una variabile conteggio. È possibile creare nel nuovo file una variabile conteggio, che contiene il numero delle righe dei dati correnti utilizzate per creare una riga nel nuovo file dati.

Creazione di variabili indicatore. Le variabili indici possono essere utilizzate per creare **variabili indicatore** nel nuovo file di dati. È possibile creare una nuova variabile per ogni valore univoco della variabile indice. Le variabili indicatore indicano la presenza o l'assenza di un valore per un caso. Il valore della variabile indicatore è 1 se il caso contiene un valore, altrimenti è 0.

Esempio. La variabile indice è *prodotto*. Tale variabile consente di registrare i prodotti acquistati da un cliente. I dati originali sono:

cliente	prodotto
1	pulcino
1	uova
2	uova
3	pulcino

La creazione di una variabile indicatore risulta in una variabile per ogni valore univoco di *prodotto*. I dati ristrutturati sono:

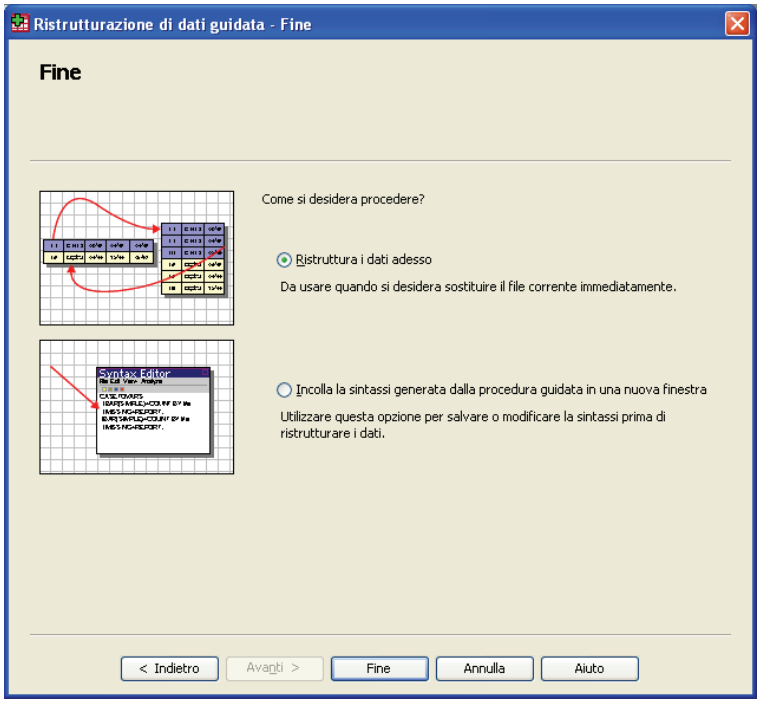
cliente	indpulcino	induova
1	1	1
2	0	1
3	1	0

In questo esempio, i dati ristrutturati potrebbero essere utilizzati per ottenere i conteggi di frequenze dei prodotti acquistati dai clienti.

Ristrutturazione di dati guidata: Fine

È il passo conclusivo della procedura. Stabilire quale operazione si desidera eseguire per le specifiche.

Figura 9-35
Ristrutturazione di dati guidata: Fine


Ristrutturazione di dati guidata - Fine
Fine
Come si desidera procedere?
Ristruttura i dati adesso
Da usare quando si desidera sostituire il file corrente immediatamente.
Incolla la sintassi generata dalla procedura guidata in una nuova finestra
Utilizzare questa opzione per salvare o modificare la sintassi prima di ristrutturare i dati.
< Indietro
Avanti >
Fine
Annulla
Aiuto

- **Ristruttura ora.** Verrà creato il nuovo file ristrutturato. Specificare se il file corrente deve essere sostituito immediatamente.

Nota: Se i dati originali sono ponderati, lo saranno anche i nuovi dati, a meno che la variabile utilizzata come peso non sia ristrutturata o esclusa dal nuovo file.

- **Incolla sintassi.** La sintassi generata viene incollata in una finestra di sintassi. Selezionare questa opzione se non si è certi di voler sostituire il file corrente oppure se si desidera modificare la sintassi o salvarla per modificarla in seguito.

Capitolo
10

Utilizzo dell'output

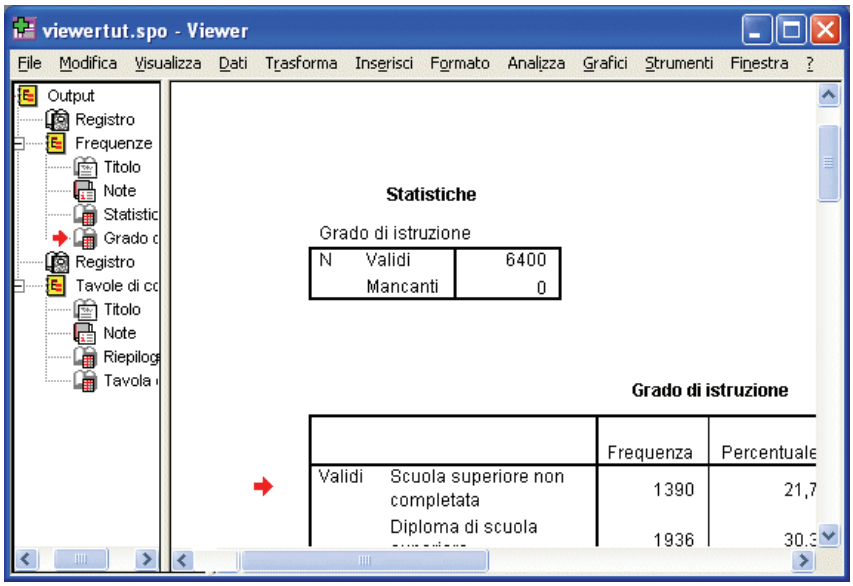
Quando si esegue una procedura, i risultati vengono visualizzati in una finestra denominata Viewer. In questa finestra è possibile spostarsi rapidamente verso qualsiasi parte dell'output che si desidera visualizzare. È inoltre possibile gestire l'output e creare un documento che contenga esattamente l'output desiderato.

Viewer

I risultati vengono visualizzati nel Viewer. È possibile utilizzare il Viewer per:

- Spostarsi nei risultati.
- Visualizzare o nascondere le tabelle e i grafici selezionati.
- Modificare l'ordine di visualizzazione dell'output spostando gli elementi selezionati.
- Spostare elementi tra il Viewer e altre applicazioni.

Figura 10-1
Viewer



Il Viewer è suddiviso in due riquadri:

- Il riquadro sinistro contiene una vista riassuntiva del contenuto dell'output.
- Il riquadro destro contiene tabelle statistiche, grafici e output testuale.

È possibile fare clic su un elemento nella vista riassuntiva per passare direttamente alla tabella o al grafico corrispondente. È possibile fare clic e trascinare il bordo destro della vista riassuntiva per modificarne la larghezza.

Attivazione e disattivazione della visualizzazione dell'output

Nel Viewer è possibile visualizzare o nascondere in modo selettivo singole tabelle o risultati inclusi in una procedura. Questo processo risulta utile se si desidera visualizzare solo una parte dell'output.

Per nascondere tabelle e grafici

- Fare doppio clic sull'icona a forma di libro dell'elemento corrispondente nella vista riassuntiva del Viewer.

  o

- Fare clic sull'elemento per selezionarlo.
- Dai menu, scegliere:
  Visualizza
  Nascondi

  o

- Fare clic sull'icona a forma di libro chiuso (Nascondi) sulla barra degli strumenti.

  Verrà attivata l'icona a forma di libro aperto (Visualizza), ad indicare che l'elemento è nascosto.

Per nascondere i risultati della procedura

- Fare clic sulla casella a sinistra del nome della procedura nella vista riassuntiva.

  Verranno nascosti tutti i risultati della procedura e la vista riassuntiva risulterà compressa.

Spostamento, copia ed eliminazione dell'output

È possibile riorganizzare l'output copiando, spostando o eliminando un elemento o un gruppo di elementi.

Per spostare l'output nel Viewer

- Selezionare gli elementi nella vista riassuntiva o nel riquadro del contenuto.
- Trascinare gli elementi selezionati in un'altra posizione.

Per eliminare l'output dal Viewer

- Selezionare gli elementi nella vista riassuntiva o nel riquadro del contenuto.

- Premere il tasto Canc.

o

- Dai menu, scegliere:

Modifica
  Elimina

Modifica dell'allineamento iniziale

Per impostazione predefinita, inizialmente tutto l'output è allineato a sinistra. Per modificare l'allineamento iniziale dei nuovi elementi di output:

- Dai menu, scegliere:

Modifica
  Opzioni

- Fare clic sulla scheda Viewer.
- Nel gruppo Stato iniziale dell'output, selezionare il tipo di elemento (ad esempio tabella pivot, grafico, output testuale).
- Selezionare l'opzione di allineamento desiderata.

Modifica dell'allineamento degli elementi dell'output

- Nella vista riassuntiva o nel riquadro del contenuto, selezionare gli elementi che si desidera allineare.
- Dai menu, scegliere:

Formato
  Allinea a sinistra

o

Formato
  Centra

o

Formato
  Allinea a destra

Vista riassuntiva del Viewer

La vista riassuntiva include un sommario del documento del Viewer. È possibile utilizzare la vista riassuntiva per spostarsi nell'output ed impostarne la visualizzazione. La maggior parte delle operazioni eseguite nella vista riassuntiva vengono applicate all'output.

- Gli elementi selezionati nella vista riassuntiva verranno visualizzati anche nell'output.
- Gli elementi spostati nella vista riassuntiva verranno spostati anche nell'output.
- Se la vista riassuntiva viene compressa, l'output di tutti gli elementi ai livelli compressi verrà nascosto.

Impostazione della visualizzazione dell'output. Per impostare le modalità di visualizzazione dell'output è possibile:

- Espandere e comprimere la vista riassuntiva.
- Modificare il livello degli elementi selezionati nella vista riassuntiva.
- Modificare le dimensioni degli elementi nella vista riassuntiva.
- Modificare il tipo di carattere utilizzato nella vista riassuntiva.

Per comprimere ed espandere la vista riassuntiva

▶ Fare clic sulla casella a sinistra dell'elemento che si desidera comprimere o espandere nella vista riassuntiva.

o

▶ Fare clic sull'elemento nella vista riassuntiva.

▶ Dai menu, scegliere:

Visualizza
  Comprimi

o

Visualizza
  Espandi

Per modificare il livello degli elementi

▶ Fare clic sull'elemento nella vista riassuntiva.

▶ Fare clic sulla freccia rivolta verso sinistra sulla barra degli strumenti della vista riassuntiva per spostare l'elemento avanti di un livello (a sinistra).

o

Fare clic sulla freccia rivolta verso destra sulla barra degli strumenti della vista riassuntiva per spostare l'elemento indietro di un livello (a destra).

o

▶ Dai menu, scegliere:

Modifica
  Vista riassuntiva
    Promuovi

o

Modifica
  Vista riassuntiva
    Retrocedi

La modifica del livello risulta particolarmente utile dopo aver spostato gli elementi nella vista riassuntiva. Spostando alcuni elementi è possibile modificare il livello degli elementi nella vista riassuntiva. Per ripristinare i livelli originali, utilizzare i pulsanti freccia destra e sinistra sulla barra degli strumenti.

Per modificare le dimensioni degli elementi nella vista riassuntiva

- Dai menu, scegliere:
  Visualizza
  Dimensioni vista riassuntiva
- Selezionare le dimensioni della vista riassuntiva (Piccola, Media o Grande).

Per modificare il carattere nella vista riassuntiva

- Dai menu, scegliere:
  Visualizza
  Carattere vista riassuntiva...
- Selezionare un carattere.

Aggiunta di elementi nel Viewer

Nel Viewer è possibile inserire elementi quali titoli, nuovo testo, grafici o materiali di altre applicazioni.

Per aggiungere un nuovo titolo o elemento testuale al contenuto del Viewer

Nel Viewer è possibile inserire elementi testuali che non siano collegati a una tabella o a un grafico.

- Fare clic sulla tabella, sul grafico o sull'oggetto che precederà il titolo o il testo.
- Dai menu, scegliere:
  Inserisci
  Nuovo titolo

  o

  Inserisci
  Nuovo testo
- Fare doppio clic sul nuovo oggetto.
- Immettere il testo.

Per aggiungere un file di testo

- Nell'output o nella vista riassuntiva del Viewer, fare clic sulla tabella, sul grafico o sull'oggetto che precederà il testo.
- Dai menu, scegliere:
  Inserisci
  File di testo...
- Selezionare un file di testo.

  Fare doppio clic sul testo da modificare.

Per incollare oggetti nel Viewer

Nel Viewer possono essere incollati oggetti di altre applicazioni. È possibile utilizzare i comandi Incolla dopo o Incolla speciale. Entrambi i comandi consentono di incollare il nuovo oggetto dopo l'oggetto selezionato nel Viewer. Utilizzare il comando Incolla speciale per scegliere il formato dell'oggetto incollato.

Individuazione e sostituzione di informazioni nel Viewer

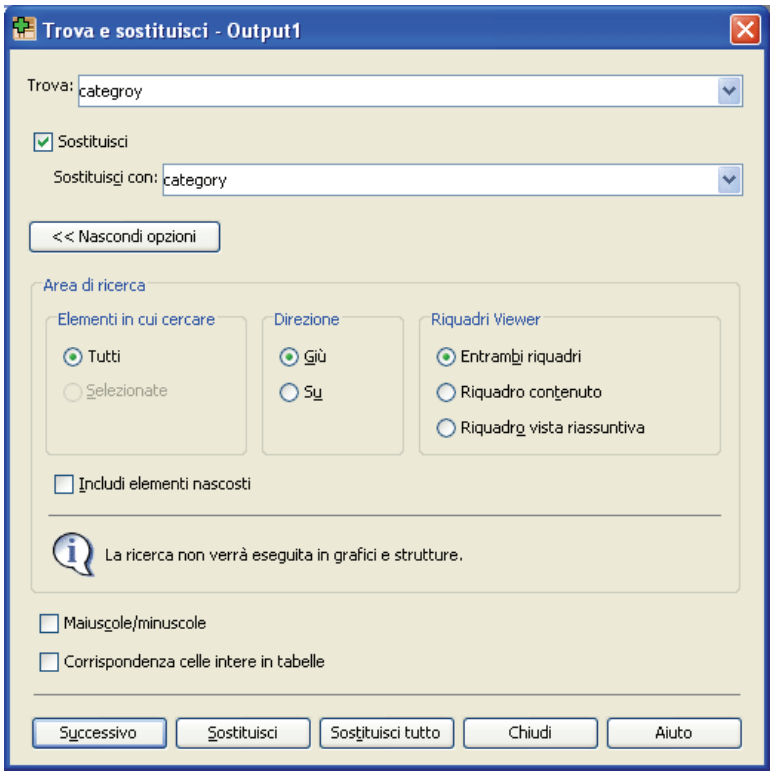
- Per trovare o sostituire informazioni nel Viewer, dai menu, scegliere:

Modifica
  Trova

o

Modifica
  Sostituisci

Figura 10-2
Finestra di dialogo Trova e sostituisci



È possibile utilizzare Trova e sostituisci per:

- Ricercare l'intero documento o soltanto gli elementi selezionati.
- Ricercare verso il basso o verso l'alto rispetto alla posizione corrente.
- Ricercare sia nel riquadro del contenuto che nella vista riassuntiva o limitare la ricerca a uno solo di questi elementi.

- Ricercare elementi nascosti. Tali elementi includono qualsiasi elemento nascosto presente nel riquadro del contenuto (ad esempio le tabelle Note che sono nascoste per impostazione predefinita) e le righe e le colonne nascoste nelle tabelle pivot.
- Limitare i criteri di ricerca alle corrispondenze con distinzione tra le maiuscole e le minuscole.
- Limitare i criteri di ricerca nelle tabelle pivot alle corrispondenze dell'intero contenuto delle celle.

Elementi nascosti e strati delle tabelle pivot

- Gli strati che si trovano al di sotto dello strato visibile di una tabella pivot multidimensionale non vengono considerati come nascosti e verranno inclusi nell'area di ricerca anche quando gli elementi nascosti non vengono inclusi nella ricerca.
- Gli elementi nascosti includono gli elementi nascosti che si trovano nel riquadro del contenuto (ovvero contrassegnati da un'icona a forma di libro chiuso nella vista riassuntiva o inclusi nei blocchi compressi della vista riassuntiva) e le righe e le colonne delle tabelle pivot nascoste per impostazione predefinita (ad esempio le righe e le colonne vuote) o nascoste manualmente modificando la tabella e nascondendo selettivamente righe o colonne specifiche. Gli elementi nascosti non vengono inclusi nella ricerca se si seleziona esplicitamente Includi elementi nascosti.
- In entrambi i casi, l'elemento nascosto o non visibile che contiene il valore o il testo della ricerca viene visualizzato quando viene trovato, ma torna subito dopo al suo stato originale.

Copia dell'output in altre applicazioni

Gli oggetti di output possono essere copiati e incollati in altre applicazioni, ad esempio in elaboratori di testo o fogli elettronici. È possibile incollare l'output in molti formati. A seconda dell'applicazione di destinazione, è possibile che siano disponibili tutti i formati seguenti o solo una parte:

Illustrazioni (metafile). È possibile incollare le tabelle pivot e i grafici come illustrazioni metafile. È possibile ridimensionare il formato dell'immagine nell'altra applicazione. A volte è possibile eseguire alcune modifiche utilizzando gli strumenti disponibili nell'altra applicazione. Le tabelle pivot incollate come illustrazioni mantengono le caratteristiche dei bordi e del carattere. Per le tabelle pivot, selezionare File dal menu Incolla speciale. Questo formato è disponibile solo nei sistemi operativi Windows.

Formato RTF (Rich Text Format). È possibile incollare le tabelle pivot in formato RTF in altre applicazioni. Nella maggior parte delle applicazioni, la tabella pivot viene incollata come una normale tabella e può essere quindi modificata.

Nota: alcune tabelle molto ampie non vengono visualizzate correttamente in Microsoft Word.

Bitmap. È possibile incollare i grafici e altre immagini in altre applicazioni come bitmap.

BIFF. È possibile incollare il contenuto di una tabella in un foglio elettronico mantenendo la precisione numerica. Questo formato è disponibile solo nei sistemi operativi Windows.

Testo. È possibile copiare il contenuto di una tabella e incollarlo come testo. Questo processo può essere utile in applicazioni come la posta elettronica che possono ricevere o trasmettere solo testo.

Se l'applicazione di destinazione supporta più formati disponibili, è possibile che contenga una voce di menu Incolla speciale che consenta di selezionare il formato desiderato o che mostri automaticamente un elenco dei formati disponibili.

Per copiare e incollare gli elementi di output in un'altra applicazione

- ▶ Selezionare gli oggetti nella vista riassuntiva o nel riquadro del contenuto del Viewer.
- ▶ Dai menu del Viewer scegliere:

  Modifica
  Copia
- ▶ Dai menu dell'applicazione di destinazione, scegliere:

  Modifica
  Incolla

  o

  Modifica
  Incolla speciale...

  - **Incolla.** L'output viene copiato negli Appunti in diversi formati. Ciascuna applicazione determina il formato migliore da utilizzare con il comando Incolla.
  - **Incolla speciale.** L'output viene copiato negli Appunti in diversi formati. Il comando Incolla speciale consente di selezionare il formato desiderato dall'elenco dei formati disponibili nell'applicazione di destinazione.

Copiare e incollare viste modelli. Tutte le viste modelli, comprese le tabelle, vengono copiate come immagini grafiche.

Copiare e incollare tabelle pivot. Quando si incollano le tabelle pivot in formato Word/RTF, le tabelle troppo grandi per la larghezza del documento verranno mandate a capo, ridotte di dimensioni per adattarle alla larghezza del documento oppure lasciate invariate, a seconda delle impostazioni delle tabelle pivot. Per ulteriori informazioni, vedere Opzioni tabella pivot in Capitolo 48 a pag. 534.

Esporta output

Il comando Export Output salva l'output del Viewer nei formati HTML, testo, Word/RTF, Excel, PowerPoint (richiede PowerPoint 97 o versione successiva) e PDF. Anche i grafici possono essere esportati in una vasta gamma di formati.

Nota: l'esportazione in PowerPoint è disponibile solo nei sistemi operativi Windows e non nella versione per studenti.

Per esportare l'output

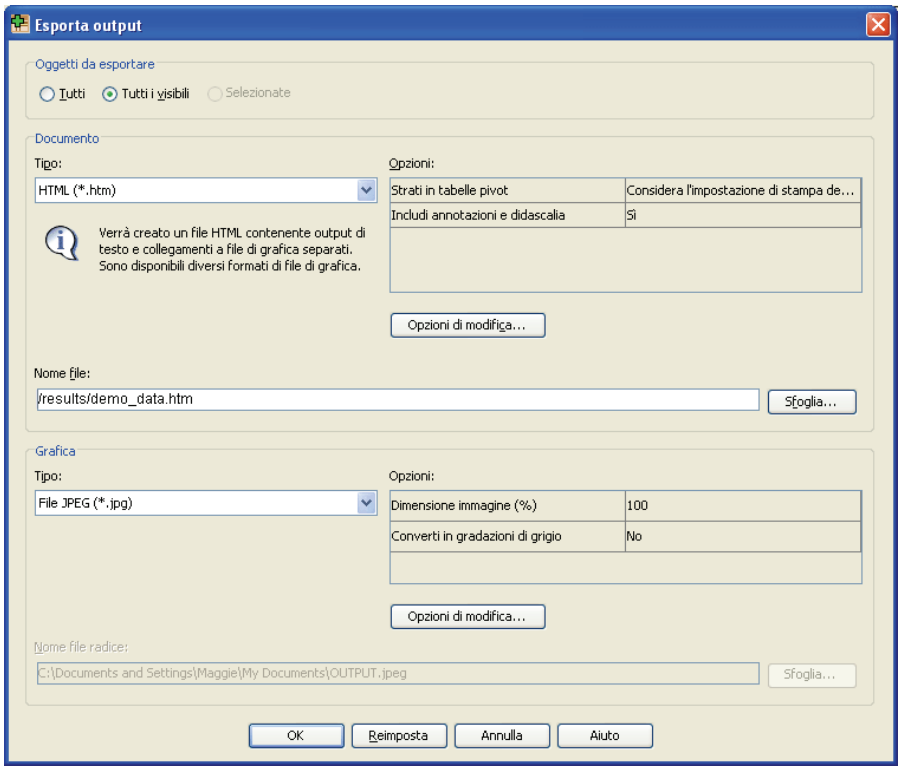
- ▶ Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).

▶ Dai menu, scegliere:

File
  Esporta...

▶ Specificare un nome di file (o il prefisso di un grafico) e selezionare un formato di esportazione.

Figura 10-3
Finestra di dialogo Esporta output



Oggetti da esportare. È possibile esportare tutti gli oggetti nel Viewer, tutti gli oggetti visibili oppure solo gli oggetti selezionati.

Tipo documento. Sono disponibili le seguenti opzioni:

- **Word/RTF (*.doc).** Le tabelle pivot vengono esportate come tabelle Word con tutti gli attributi di formattazione; ad esempio bordi delle celle, stili del carattere e colori dello sfondo. L'output testuale viene esportato come HTML formattato. I grafici, i diagrammi ad albero e le viste del modello vengono inclusi nel formato PNG.

  Nota: alcune tabelle molto ampie non vengono visualizzate correttamente in Microsoft Word.

- **Excel (*.xls).** Le righe, le colonne e le celle delle tabelle pivot vengono esportate come righe, colonne e celle di Excel, con tutti gli attributi di formattazione, ad esempio, con i bordi delle celle, gli stili del carattere, il colore di sfondo delle celle e così via. L'output testuale viene esportato con tutti gli attributi del carattere. Ogni riga nell'output testuale è una riga del file di Excel, con tutti i contenuti di una riga raccolti in una singola cella. I grafici, i diagrammi ad albero e le viste del modello vengono inclusi nel formato PNG.

- **HTML (*.htm).** Le tabelle pivot vengono esportate come tabelle HTML. L'output testuale viene esportato come HTML preformattato. I grafici, i diagrammi ad albero e le viste dei modelli vengono incorporati per riferimento e possono essere esportati in un formato adatto all'inserimento in documenti HTML (ad esempio, PNG e JPEG).
- **Portable Document Format (*.pdf).** Tutto l'output viene esportato così come appare in Anteprima di stampa con tutti gli attributi di formattazione invariati.
- **File PowerPoint (*.ppt).** Le tabelle pivot vengono esportate come tabelle Word ed incorporate in diapositive diverse del file PowerPoint; una diapositiva per ciascuna tabella pivot. Tutti gli attributi di formattazione della tabella pivot (quali i bordi delle celle, gli stili dei caratteri, i colori di sfondo, ecc.) vengono mantenuti. I grafici, i diagrammi ad albero e le viste del modello vengono esportati in formato TIFF. L'output testuale non è incluso.

  L'esportazione in PowerPoint è disponibile solo nei sistemi operativi Windows.
- **Testo (*.txt).** I formati di output testuale includono testo normale, UTF-8 e UTF-16. Le tabelle pivot possono essere esportate in formato delimitato da tabulazioni o da spazi. Tutto l'output testuale viene esportato in formato delimitato da spazi. Per i grafici, i diagrammi ad albero e le viste del modello viene inserita una riga nel file di testo per ogni immagine, che indica il nome del file di immagine.
- **Nessuno (solo grafica).** Sono disponibili diversi formati di esportazione: EPS, JPEG, TIFF, PNG e BMP. Nei sistemi operativi Windows è disponibile anche il formato EMF (Metafile avanzato).

Sistema di gestione dell'output. È possibile esportare automaticamente tutto l'output oppure solo i tipi di output specificati dall'utente in file di dati in formato Word, Excel, PDF, HTML, testo e SPSS Statistics. Per ulteriori informazioni, vedere Sistema di gestione dell'output in Capitolo 52 a pag. 585.

Opzioni HTML

Le seguenti opzioni sono disponibili per l'esportazione dell'output in formato HTML:

Strati in tabelle pivot. Per impostazione predefinita, l'inclusione o l'esclusione degli strati nelle tabelle pivot è controllata dalle proprietà delle tabelle per ogni tabella pivot. È possibile ignorare questa impostazione e includere tutti gli strati o escludere tutto tranne lo strato visibile al momento. Per ulteriori informazioni, vedere Proprietà tabella: Stampa in Capitolo 11 a pag. 255.

Includi annotazioni e didascalie. Controlla l'inclusione o l'esclusione di tutte le annotazioni e didascalie delle tabelle pivot.

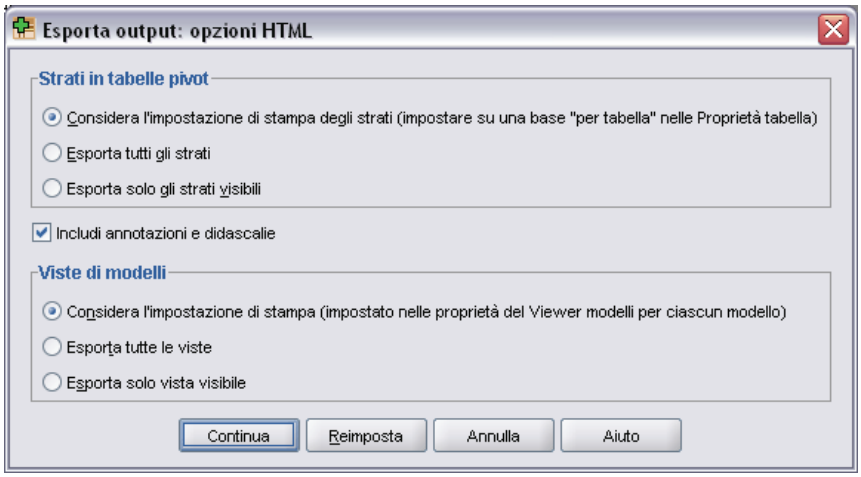
Viste dei modelli. Per impostazione predefinita, l'inclusione o l'esclusione delle viste dei modelli è controllata dalle proprietà di ogni modello. È possibile ignorare questa impostazione e includere tutte le viste o escludere tutto tranne la vista attualmente visibile. Per ulteriori informazioni, vedere Proprietà dei modelli in Capitolo 12 a pag. 266. Nota: tutte le viste dei modelli, comprese le tabelle, vengono esportate come immagini.

Nota: per il formato HTML, è possibile controllare anche il formato dei file immagine per i grafici esportati. Per ulteriori informazioni, vedere Opzioni di formato grafico a pag. 234.

Per impostare opzioni di importazione HTML

- Selezionare HTML come formato di esportazione.
- Fare clic su Opzioni di modifica.

Figura 10-4
Opzioni di esportazione dell'output in HTML


Esporta output: opzioni HTML
Strati in tabelle pivot
Considera l'impostazione di stampa degli strati (impostare su una base "per tabella" nelle Proprietà tabella)
Esporta tutti gli strati
Esporta solo gli strati visibili
Includi annotazioni e didascalie
Viste di modelli
Considera l'impostazione di stampa (impostato nelle proprietà del Viewer modelli per ciascun modello)
Esporta tutte le viste
Esporta solo vista visibile
Continua
Reimposta
Annulla
Aiuto

Opzioni Word/RTF

Le seguenti opzioni sono disponibili per l'esportazione dell'output in formato Word:

Strati in tabelle pivot. Per impostazione predefinita, l'inclusione o l'esclusione degli strati nelle tabelle pivot è controllata dalle proprietà delle tabelle per ogni tabella pivot. È possibile ignorare questa impostazione e includere tutti gli strati o escludere tutto tranne lo strato visibile al momento. Per ulteriori informazioni, vedere Proprietà tabella: Stampa in Capitolo 11 a pag. 255.

Tabelle pivot grandi. Controlla il trattamento delle tabelle troppo grandi per la larghezza definita del documento. Per impostazione predefinita, la tabella viene mandata a capo. La tabella viene divisa in sezioni e le etichette di riga vengono ripetute per ogni sezione. In alternativa è possibile ridurre le tabelle grandi oppure non apportare alcuna modifica e consentire alle tabelle grandi di estendersi oltre la larghezza definita del documento.

Includi annotazioni e didascalie. Controlla l'inclusione o l'esclusione di tutte le annotazioni e didascalie delle tabelle pivot.

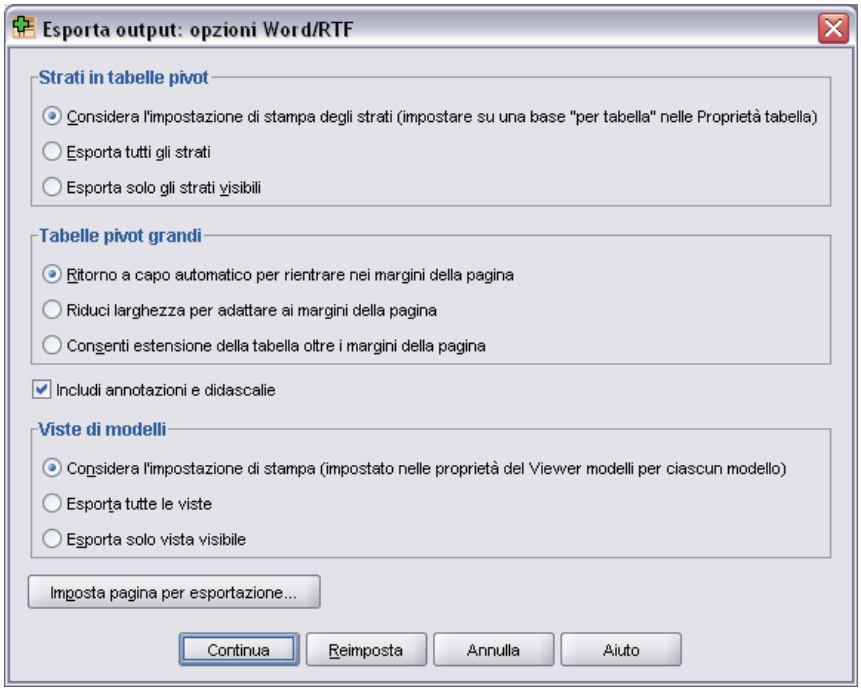
Viste dei modelli. Per impostazione predefinita, l'inclusione o l'esclusione delle viste dei modelli è controllata dalle proprietà di ogni modello. È possibile ignorare questa impostazione e includere tutte le viste o escludere tutto tranne la vista attualmente visibile. Per ulteriori informazioni, vedere Proprietà dei modelli in Capitolo 12 a pag. 266. Nota: tutte le viste dei modelli, comprese le tabelle, vengono esportate come immagini.

Imposta pagina per esportazione. Questa opzione apre una finestra di dialogo in cui è possibile definire le dimensioni e i margini della pagina per il documento esportato. La larghezza del documento utilizzata per determinare se andare a capo o ridurre le dimensioni è data dalla larghezza della pagina meno i margini sinistro e destro.

Per impostare le opzioni di esportazione in Word

- Selezionare Word/RTF come formato di esportazione.
- Fare clic su Opzioni di modifica.

Figura 10-5
Opzioni di esportazione dell'output in Word


Esporta output: opzioni Word/RTF
Strati in tabelle pivot
Considera l'impostazione di stampa degli strati (impostare su una base "per tabella" nelle Proprietà tabella)
Esporta tutti gli strati
Esporta solo gli strati visibili
Tabelle pivot grandi
Ritorno a capo automatico per rientrare nei margini della pagina
Riduci larghezza per adattare ai margini della pagina
Consenti estensione della tabella oltre i margini della pagina
Includi annotazioni e didascalie
Viste di modelli
Considera l'impostazione di stampa (impostato nelle proprietà del Viewer modelli per ciascun modello)
Esporta tutte le viste
Esporta solo vista visibile
Imposta pagina per esportazione...
Continua
Reimposta
Annulla
Aiuto

Opzioni Excel

Le seguenti opzioni sono disponibili per l'esportazione dell'output in formato Excel:

Creare un foglio o una cartella di lavoro, oppure modificare un foglio di lavoro esistente. Per impostazione predefinita, viene creata una nuova cartella di lavoro. Se esiste già un file con il nome specificato, questo verrà sovrascritto. Se si seleziona l'opzione per creare un foglio di lavoro ed esiste già un foglio di lavoro con il nome specificato nel file, questo verrà sovrascritto. Se si seleziona l'opzione di modifica di un foglio di lavoro esistente, è necessario specificare anche il nome del foglio di lavoro (questo è facoltativo se si crea un foglio di lavoro). I nomi dei fogli di lavoro non possono superare i 31 caratteri e non possono contenere barre o barre rovesciate, parentesi quadre, punti di domanda o asterischi.

Se si modifica un foglio di lavoro esistente, i grafici, le viste modello e i diagrammi ad albero non sono inclusi nell'output esportato.

Posizione in foglio di lavoro. Controlla la posizione all'interno del foglio di lavoro dell'output esportato. Per impostazione predefinita, l'output esportato verrà aggiunto all'ultima colonna che presenta del contenuto, iniziando dalla prima riga e senza modificare eventuali contenuti esistenti. Si tratta di un buon metodo per aggiungere nuove colonne a un foglio di lavoro esistente. L'aggiunta di un output esportato dopo l'ultima riga è un buon modo per aggiungere nuove righe a un foglio di lavoro esistente. L'aggiunta di un output esportato in corrispondenza di una cella specifica sovrascriverà qualsiasi contenuto esistente presente nell'area in cui viene aggiunto tale output esportato.

Strati in tabelle pivot. Per impostazione predefinita, l'inclusione o l'esclusione degli strati nelle tabelle pivot è controllata dalle proprietà delle tabelle per ogni tabella pivot. È possibile ignorare questa impostazione e includere tutti gli strati o escludere tutto tranne lo strato visibile al momento. Per ulteriori informazioni, vedere Proprietà tabella: Stampa in Capitolo 11 a pag. 255.

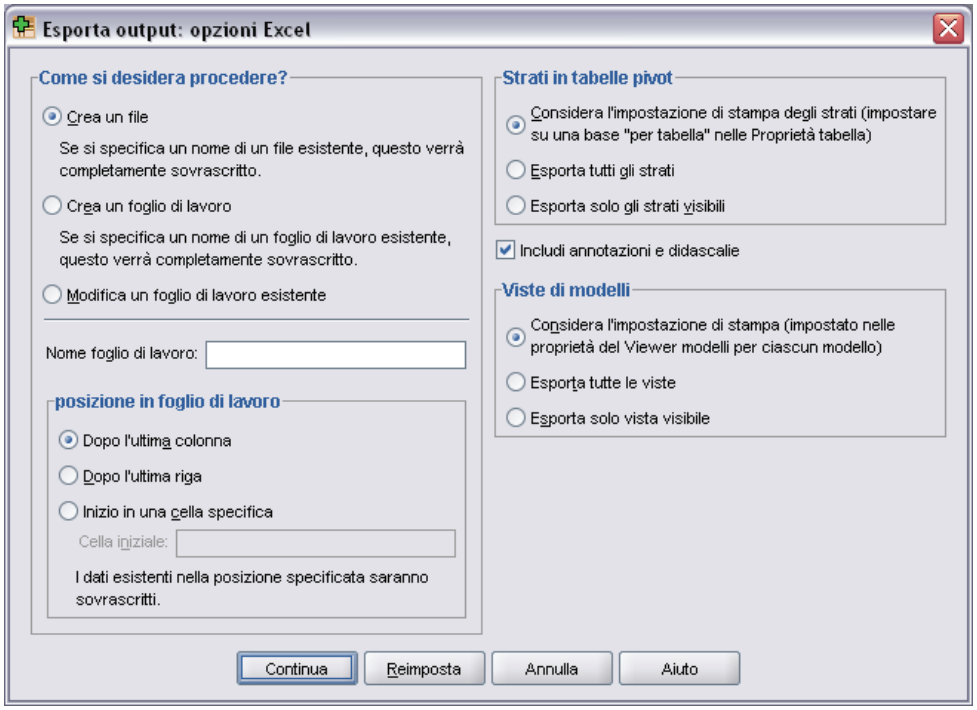
Includi annotazioni e didascalie. Controlla l'inclusione o l'esclusione di tutte le annotazioni e didascalie delle tabelle pivot.

Viste dei modelli. Per impostazione predefinita, l'inclusione o l'esclusione delle viste dei modelli è controllata dalle proprietà di ogni modello. È possibile ignorare questa impostazione e includere tutte le viste o escludere tutto tranne la vista attualmente visibile. Per ulteriori informazioni, vedere Proprietà dei modelli in Capitolo 12 a pag. 266. Nota: tutte le viste dei modelli, comprese le tabelle, vengono esportate come immagini.

Per impostare le opzioni di eportazione in Excel

- Selezionare Excel come formato di esportazione.
- Fare clic su Opzioni di modifica.

Figura 10-6
Opzioni di esportazione dell'output in Excel


Esporta output: opzioni Excel
Come si desidera procedere?
Crea un file
Se si specifica un nome di un file esistente, questo verrà completamente sovrascritto.
Crea un foglio di lavoro
Se si specifica un nome di un foglio di lavoro esistente, questo verrà completamente sovrascritto.
Modifica un foglio di lavoro esistente
Nome foglio di lavoro:
posizione in foglio di lavoro
Dopo l'ultima colonna
Dopo l'ultima riga
Inizio in una cella specifica
Cella iniziale:
I dati esistenti nella posizione specificata saranno sovrascritti.
Strati in tabelle pivot
Considera l'impostazione di stampa degli strati (impostare su una base "per tabella" nelle Proprietà tabella)
Esporta tutti gli strati
Esporta solo gli strati visibili
Includi annotazioni e didascalie
Viste di modelli
Considera l'impostazione di stampa (impostato nelle proprietà del Viewer modelli per ciascun modello)
Esporta tutte le viste
Esporta solo vista visibile
Continua
Reimposta
Annulla
Aiuto

Opzioni di PowerPoint

Per PowerPoint sono disponibili le seguenti opzioni:

Strati in tabelle pivot. Per impostazione predefinita, l'inclusione o l'esclusione degli strati nelle tabelle pivot è controllata dalle proprietà delle tabelle per ogni tabella pivot. È possibile ignorare questa impostazione e includere tutti gli strati o escludere tutto tranne lo strato visibile al momento. Per ulteriori informazioni, vedere Proprietà tabella: Stampa in Capitolo 11 a pag. 255.

Tabelle pivot grandi. Controlla il trattamento delle tabelle troppo grandi per la larghezza definita del documento. Per impostazione predefinita, la tabella viene mandata a capo. La tabella viene divisa in sezioni e le etichette di riga vengono ripetute per ogni sezione. In alternativa è possibile ridurre le tabelle grandi oppure non apportare alcuna modifica e consentire alle tabelle grandi di estendersi oltre la larghezza definita del documento.

Includi annotazioni e didascalie. Controlla l'inclusione o l'esclusione di tutte le annotazioni e didascalie delle tabelle pivot.

Usa le voci della vista riassuntiva di Viewer come titoli di diapositiva. Include un titolo in ciascuna diapositiva creata con l'esportazione. Ciascuna diapositiva contiene un'unica voce esportata da Viewer. Il titolo è costituito dalla voce corrispondente all'elemento desiderato nella vista riassuntiva di Viewer.

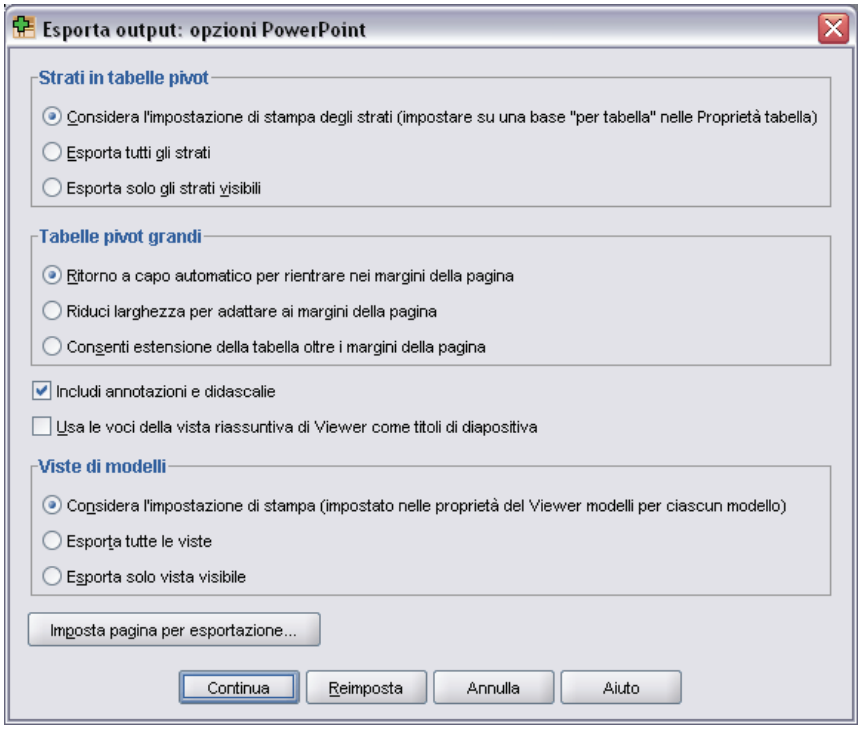
Viste dei modelli. Per impostazione predefinita, l'inclusione o l'esclusione delle viste dei modelli è controllata dalle proprietà di ogni modello. È possibile ignorare questa impostazione e includere tutte le viste o escludere tutto tranne la vista attualmente visibile. Per ulteriori informazioni, vedere Proprietà dei modelli in Capitolo 12 a pag. 266. Nota: tutte le viste dei modelli, comprese le tabelle, vengono esportate come immagini.

Imposta pagina per esportazione. Questa opzione apre una finestra di dialogo in cui è possibile definire le dimensioni e i margini della pagina per il documento esportato. La larghezza del documento utilizzata per determinare se andare a capo o ridurre le dimensioni è data dalla larghezza della pagina meno i margini sinistro e destro.

Per impostare le opzioni di Esportazione in PowerPoint

- Selezionare PowerPoint come formato di esportazione.
- Fare clic su Opzioni di modifica.

Figura 10-7
Finestra di dialogo Opzioni di esportazione in PowerPoint


Esporta output: opzioni PowerPoint
Strati in tabelle pivot
Considera l'impostazione di stampa degli strati (impostare su una base "per tabella" nelle Proprietà tabella)
Esporta tutti gli strati
Esporta solo gli strati visibili
Tabelle pivot grandi
Ritorno a capo automatico per rientrare nei margini della pagina
Riduci larghezza per adattare ai margini della pagina
Consenti estensione della tabella oltre i margini della pagina
Includi annotazioni e didascalie
Usa le voci della vista riassuntiva di Viewer come titoli di diapositiva
Viste di modelli
Considera l'impostazione di stampa (impostato nelle proprietà del Viewer modelli per ciascun modello)
Esporta tutte le viste
Esporta solo vista visibile
Imposta pagina per esportazione...
Continua
Reimposta
Annulla
Aiuto

Nota: l'esportazione in PowerPoint è disponibile solo nei sistemi operativi Windows.

Opzioni PDF

Per il formato PDF sono disponibili le seguenti opzioni:

Segnalibri incorporati. Questa opzione include nel documento PDF i segnalibri corrispondenti alle voci di Vista riassuntiva del Viewer. Analogamente alla Vista riassuntiva del Viewer, i segnalibri facilitano lo spostamento nei documenti contenenti un elevato numero di oggetti di output.

Caratteri incorporati. L'incorporazione dei caratteri garantisce che il documento PDF avrà lo stesso aspetto su tutti i computer. Altrimenti, se alcuni caratteri del documento non fossero disponibili sul computer usato per visualizzare (o stampare) il documento PDF, la sostituzione dei caratteri potrebbe generare risultati deludenti.

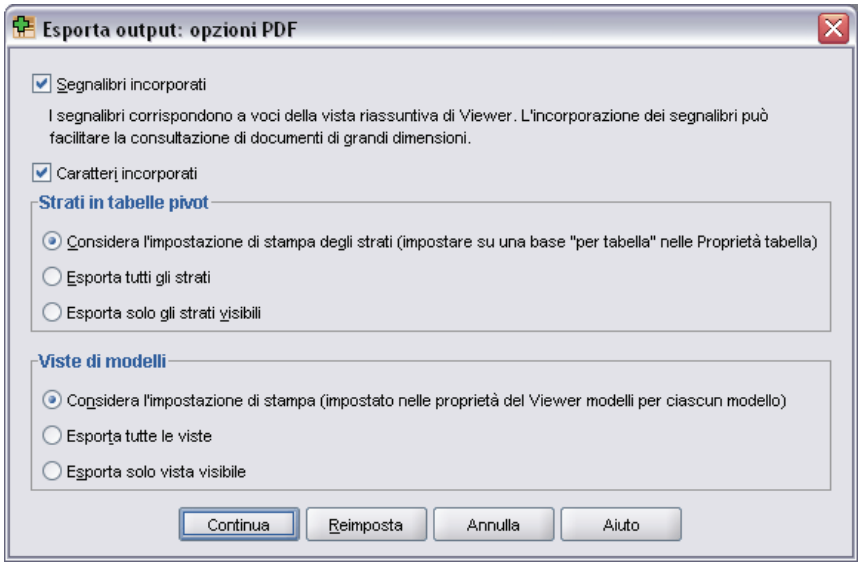
Strati in tabelle pivot. Per impostazione predefinita, l'inclusione o l'esclusione degli strati nelle tabelle pivot è controllata dalle proprietà delle tabelle per ogni tabella pivot. È possibile ignorare questa impostazione e includere tutti gli strati o escludere tutto tranne lo strato visibile al momento. Per ulteriori informazioni, vedere Proprietà tabella: Stampa in Capitolo 11 a pag. 255.

Viste dei modelli. Per impostazione predefinita, l'inclusione o l'esclusione delle viste dei modelli è controllata dalle proprietà di ogni modello. È possibile ignorare questa impostazione e includere tutte le viste o escludere tutto tranne la vista attualmente visibile. Per ulteriori informazioni, vedere Proprietà dei modelli in Capitolo 12 a pag. 266. Nota: tutte le viste dei modelli, comprese le tabelle, vengono esportate come immagini.

Per impostare le opzioni di Esportazione in PDF

- Selezionare Portable Document Format come formato di esportazione.
- Fare clic su Opzioni di modifica.

Figura 10-8
Finestra di dialogo Opzioni PDF


Esporta output: opzioni PDF
Segnalibri incorporati
I segnalibri corrispondono a voci della vista riassuntiva di Viewer. L'incorporazione dei segnalibri può facilitare la consultazione di documenti di grandi dimensioni.
Caratteri incorporati
Strati in tabelle pivot
Considera l'impostazione di stampa degli strati (impostare su una base "per tabella" nelle Proprietà tabella)
Esporta tutti gli strati
Esporta solo gli strati visibili
Viste di modelli
Considera l'impostazione di stampa (impostato nelle proprietà del Viewer modelli per ciascun modello)
Esporta tutte le viste
Esporta solo vista visibile
Continua
Reimposta
Annulla
Aiuto

Altre impostazioni incidenti sull'output PDF

Imposta pagina/Attributi di pagina. Le dimensioni della pagina, l'orientamento, i margini, il contenuto e la visualizzazione di intestazioni e piè di pagina nonché le dimensioni del grafico stampato nei documenti PDF sono controllati dall'opzione di impostazione pagina e da quella relativa agli attributi della pagina.

Proprietà tabella/Modelli di tabelle. Il ridimensionamento di tabelle grandi e/o lunghe e la stampa di strati di tabelle sono controllati dalle proprietà della tabella per ciascuna tabella. Tali proprietà possono anche essere salvate in Modelli di tabelle. Per ulteriori informazioni, vedere Proprietà tabella: Stampa in Capitolo 11 a pag. 255.

Stampante predefinita/corrente. La risoluzione (DPI) del documento PDF corrisponde alla risoluzione corrente della stampante predefinita o correntemente selezionata (modificabile in Imposta pagina). La risoluzione massima è 1200 DPI. Se l'impostazione della stampante è maggiore, la risoluzione del documento PDF sarà 1200 DPI.

Nota: la qualità di documenti ad alta risoluzione potrebbe essere scarsa se si utilizzano stampanti con risoluzioni inferiori.

Opzioni testo

Per l'esportazione del testo sono disponibili le seguenti opzioni:

Formato tabella pivot. È possibile esportare le tabelle pivot in formato delimitato da tabulazioni o da spazi. Per il formato delimitato da spazi è anche possibile controllare:

- **Larghezza colonna.** L'opzione Adatta non manda a capo il testo dei contenuti delle colonne e di conseguenza l'ampiezza di ciascuna colonna è pari a quella dell'etichetta o del valore più ampio contenuto in tale colonna. L'opzione Personalizzata imposta una larghezza massima di colonna che viene applicata a tutte le colonne della tabella; i valori che superano tale larghezza vengono mandati a capo nella riga successiva di tale colonna.
- **Carattere del bordo riga/colonna.** Queste opzioni consentono di controllare i caratteri utilizzati per creare bordi di riga e di colonna. Per disattivare la visualizzazione dei bordi di riga e di colonna, immettere spazi vuoti al posto dei valori.

Strati in tabelle pivot. Per impostazione predefinita, l'inclusione o l'esclusione degli strati nelle tabelle pivot è controllata dalle proprietà delle tabelle per ogni tabella pivot. È possibile ignorare questa impostazione e includere tutti gli strati o escludere tutto tranne lo strato visibile al momento. Per ulteriori informazioni, vedere Proprietà tabella: Stampa in Capitolo 11 a pag. 255.

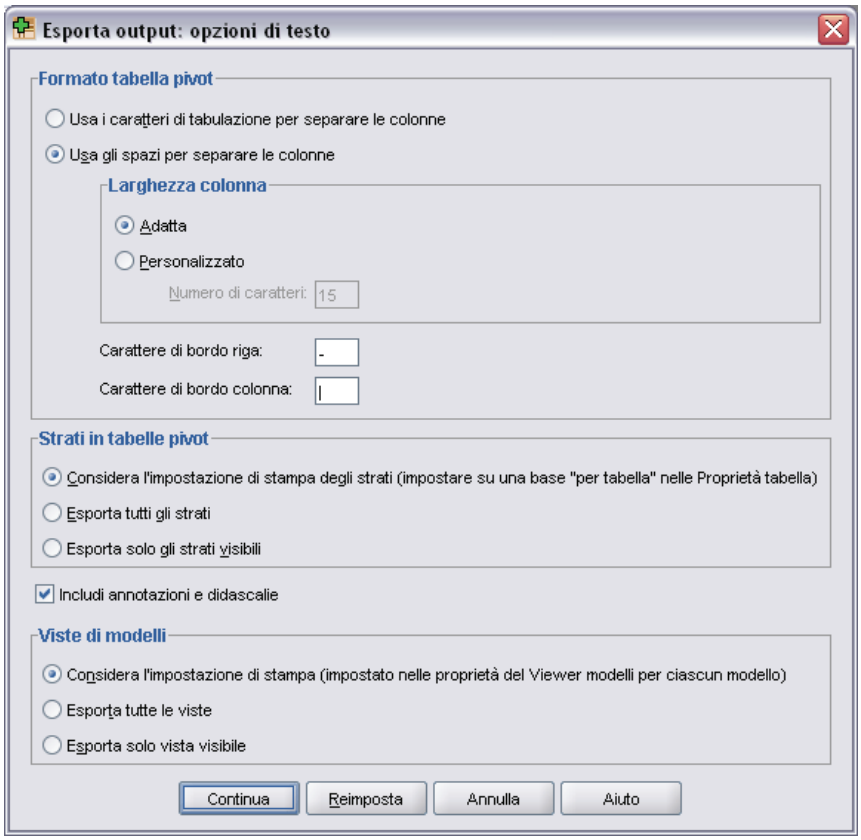
Includi annotazioni e didascalie. Controlla l'inclusione o l'esclusione di tutte le annotazioni e didascalie delle tabelle pivot.

Viste dei modelli. Per impostazione predefinita, l'inclusione o l'esclusione delle viste dei modelli è controllata dalle proprietà di ogni modello. È possibile ignorare questa impostazione e includere tutte le viste o escludere tutto tranne la vista attualmente visibile. Per ulteriori informazioni, vedere Proprietà dei modelli in Capitolo 12 a pag. 266. Nota: tutte le viste dei modelli, comprese le tabelle, vengono esportate come immagini.

Per impostare opzioni per l'esportazione di testo

- Selezionare Testo come formato di esportazione.
- Fare clic su Opzioni di modifica.

Figura 10-9
Finestra di dialogo Opzioni testo

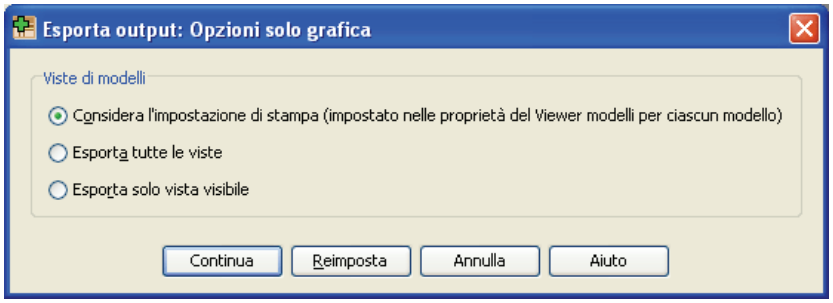

Esporta output: opzioni di testo
Formato tabella pivot
Usa i caratteri di tabulazione per separare le colonne
Usa gli spazi per separare le colonne
Larghezza colonna
Adatta
Personalizzato
Numero di caratteri: 15
Carattere di bordo riga: -
Carattere di bordo colonna: |
Strati in tabelle pivot
Considera l'impostazione di stampa degli strati (impostare su una base "per tabella" nelle Proprietà tabella)
Esporta tutti gli strati
Esporta solo gli strati visibili
Includi annotazioni e didascalie
Viste di modelli
Considera l'impostazione di stampa (impostato nelle proprietà del Viewer modelli per ciascun modello)
Esporta tutte le viste
Esporta solo vista visibile
Continua
Reimposta
Annulla
Aiuto

Opzioni solo grafica

Per l'esportazione solo della grafica sono disponibili le seguenti opzioni:

Viste dei modelli. Per impostazione predefinita, l'inclusione o l'esclusione delle viste dei modelli è controllata dalle proprietà di ogni modello. È possibile ignorare questa impostazione e includere tutte le viste o escludere tutto tranne la vista attualmente visibile. Per ulteriori informazioni, vedere Proprietà dei modelli in Capitolo 12 a pag. 266. Nota: tutte le viste dei modelli, comprese le tabelle, vengono esportate come immagini.

Figura 10-10
Opzioni solo grafica



Opzioni di formato grafico

Per i documenti HTML e in formato testo e per la sola esportazione di grafici, è possibile selezionare il formato grafico e per ogni formato grafico è possibile controllare varie impostazioni opzionali.

Per selezionare il formato grafico e le opzioni per i grafici esportati:

- Selezionare HTML, Testo o Nessuno (solo grafica) come tipo di documento.
- Selezionare il formato di file grafico dall'elenco a discesa.
- Fare clic su Opzioni di modifica per cambiare le opzioni relative al formato di file grafico selezionato.

Opzioni di esportazione dei grafici JPEG

- **Dimensione immagine.** Percentuale della dimensione del grafico originale, fino al 200%.
- **Converti in gradazioni di grigio.** Converte i colori in gradazioni di grigio.

Opzioni di esportazione dei grafici BMP

- **Dimensione immagine.** Percentuale della dimensione del grafico originale, fino al 200%.
- **Comprimi immagine per ridurre la dimensione del file.** Una tecnica di compressione senza perdita di dati che crea file più piccoli senza influire sulla qualità dell'immagine.

Opzioni di esportazione dei grafici PNG

Dimensione immagine. Percentuale della dimensione del grafico originale, fino al 200%.

Profondità colore. Consente di determinare il numero di colori nel grafico esportato. Un grafico salvato con qualsiasi profondità di colore sarà caratterizzato dal minimo del numero di colori effettivamente utilizzati e dal massimo del numero di colori consentiti dalla profondità. Ad esempio, se il grafico contiene tre colori (rosso, bianco e nero) e viene salvato con il formato a 16 colori, il grafico rimarrà a tre colori.

- Se il numero di colori nel grafico supera il numero di colori corrispondente alla profondità, verrà eseguita la retinatura dei colori per replicare quelli presenti nel grafico.
- Profondità schermo corrente è il numero di colori correntemente visualizzato sullo schermo del computer.

Opzioni di esportazione dei grafici EMF e TIFF

Dimensione immagine. Percentuale della dimensione del grafico originale, fino al 200%.

Nota: il formato EMF (Metafile avanzato) è disponibile solo nei sistemi operativi Windows.

Opzioni di esportazione dei grafici EPS

Dimensione immagine. È possibile specificare le dimensioni come percentuale delle dimensioni dell'immagine originale (fino al 200%) oppure la larghezza di un'immagine in pixel (con l'altezza determinata dal valore della larghezza e dalle proporzioni). L'immagine esportata è sempre proporzionale all'originale.

Includi immagine di anteprima TIFF. Salva un'anteprima con l'immagine EPS in formato TIFF in modo che possa essere visualizzata in applicazioni che non supportano la visualizzazione delle immagini EPS sullo schermo.

Caratteri. Controlla la modalità di gestione dei caratteri nelle immagini EPS.

- **Usa caratteri di riferimento.** Se i caratteri usati nel grafico sono disponibili per la periferica di output, vengono utilizzati questi caratteri. Altrimenti, la periferica di output utilizza altri caratteri.
- **Sostituisci caratteri con curve.** Consente di convertire i caratteri in dati di curva PostScript. Il testo non può più essere modificato come testo nelle applicazioni che consentono la modifica dei grafici EPS. Questa opzione è utile se i caratteri usati nel grafico non sono disponibili sulla periferica di output.

Stampa di documenti del Viewer

Sono disponibili due opzioni per stampare il contenuto della finestra Viewer:

Tutto l'output visibile. Consente di stampare solo gli elementi visualizzati. Gli elementi nascosti, ovvero contrassegnati da un'icona a forma di libro chiuso o nascosti negli strati compressi nella vista riassuntiva, non verranno stampati.

Selezione. Consente di stampare solo gli elementi selezionati nella vista riassuntiva e/o nell'output.

Per stampare output e grafici

▶ Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).

▶ Dai menu, scegliere:

File
  Stampa...

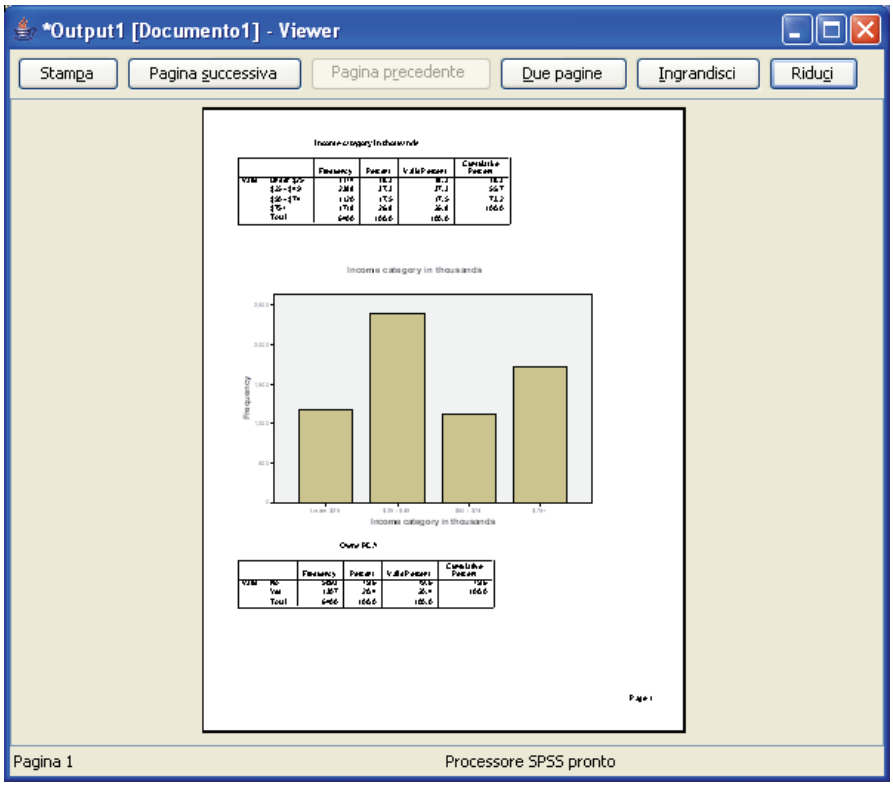
- Selezionare le impostazioni di stampa desiderate.
- Fare clic su OK per stampare.

Anteprima di stampa

Nell'Anteprima di stampa è possibile visualizzare il contenuto che verrà stampato su ogni pagina dei documenti del Viewer. Prima di stampare un documento del Viewer è consigliabile attivare l'Anteprima di stampa, in quanto vi vengono visualizzati elementi che potrebbero non risultare visibili nel Viewer, ad esempio:

- Interruzioni di pagina
- Strati nascosti di tabelle pivot
- Interruzioni in tabelle di grandi dimensioni
- Intestazioni e piè di pagina stampati in ciascuna pagina

Figura 10-11
Anteprima di stampa



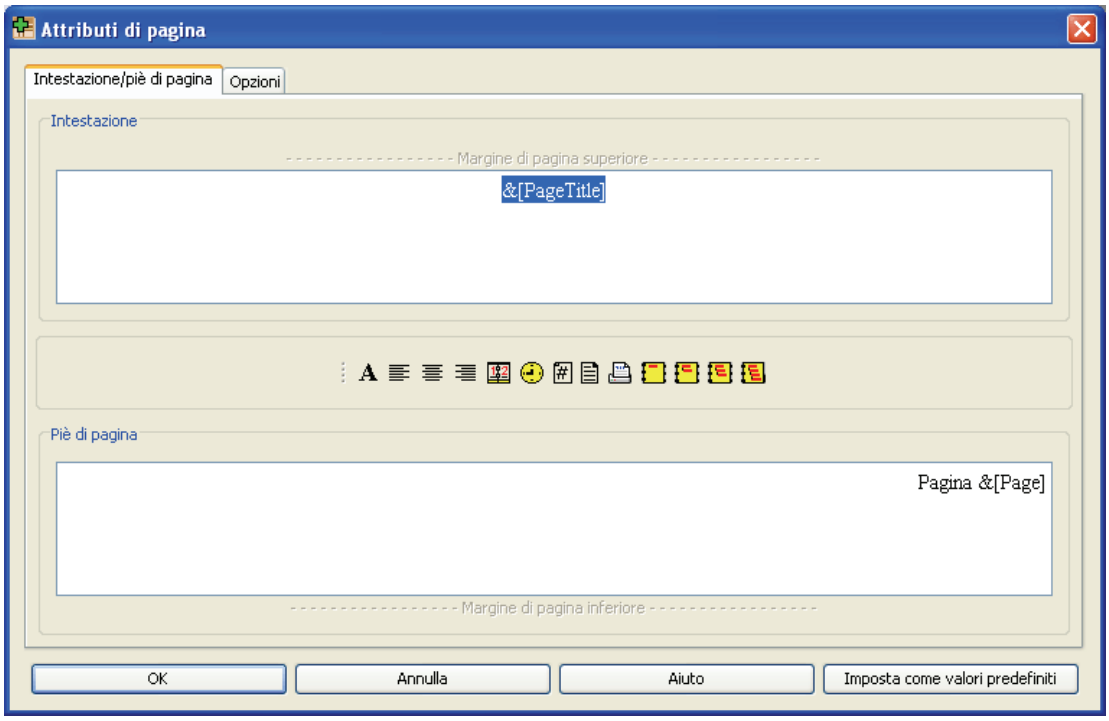
Se nel Viewer è selezionato output, nell'anteprima verrà visualizzata solo la selezione attiva. Per visualizzare in anteprima tutto l'output è necessario verificare che nel Viewer non sia stata effettuata alcuna selezione.

Attributi di pagina: intestazioni e piè di pagina

Le intestazioni e i piè di pagina sono le informazioni che vengono stampate all'inizio e alla fine di ogni pagina. È possibile specificare qualsiasi testo che si desidera utilizzare come intestazione o piè di pagina. È inoltre possibile utilizzare la barra degli strumenti al centro della finestra di dialogo per inserire:

- Data e orario
- Numeri di pagina
- Nome file Viewer
- Etichette dell'intestazione
- Titoli e sottotitoli di pagina

Figura 10-12
Finestra di dialogo Attributi di pagina, scheda Intestazione/Piè di pagina



- Imposta come valori predefiniti utilizza le impostazioni specificate qui come predefinite per i nuovi documenti Viewer. (Nota: in questo modo le impostazioni correnti sia nella scheda Intestazione/piè di pagina sia nella scheda Opzioni vengono considerate impostazioni predefinite.)
- Le etichette delle intestazioni indicano l'intestazione di primo, secondo, terzo e/o quarto livello del primo elemento di ciascuna pagina.
- Vengono stampati i titoli e i sottotitoli della pagina corrente. Questi possono essere creati con l'opzione Nuovo titolo di pagina del menu Inserisci del Viewer o con i comandi `TITLE` e `SUBTITLE`. Se non sono stati specificati titoli o sottotitoli, questa impostazione verrà ignorata.

Nota: le caratteristiche del carattere di titoli e sottotitoli della nuova pagina possono essere definite utilizzando la scheda Viewer della finestra di dialogo Opzioni (accessibile selezionando Opzioni nel menu Modifica). Le impostazioni dei caratteri di titoli e sottotitoli esistenti possono essere ridefinite modificando i titoli nel Viewer.

Per visualizzare in anteprima intestazioni e piè di pagina, selezionare Anteprima di stampa dal menu File.

Per inserire intestazioni e piè di pagina

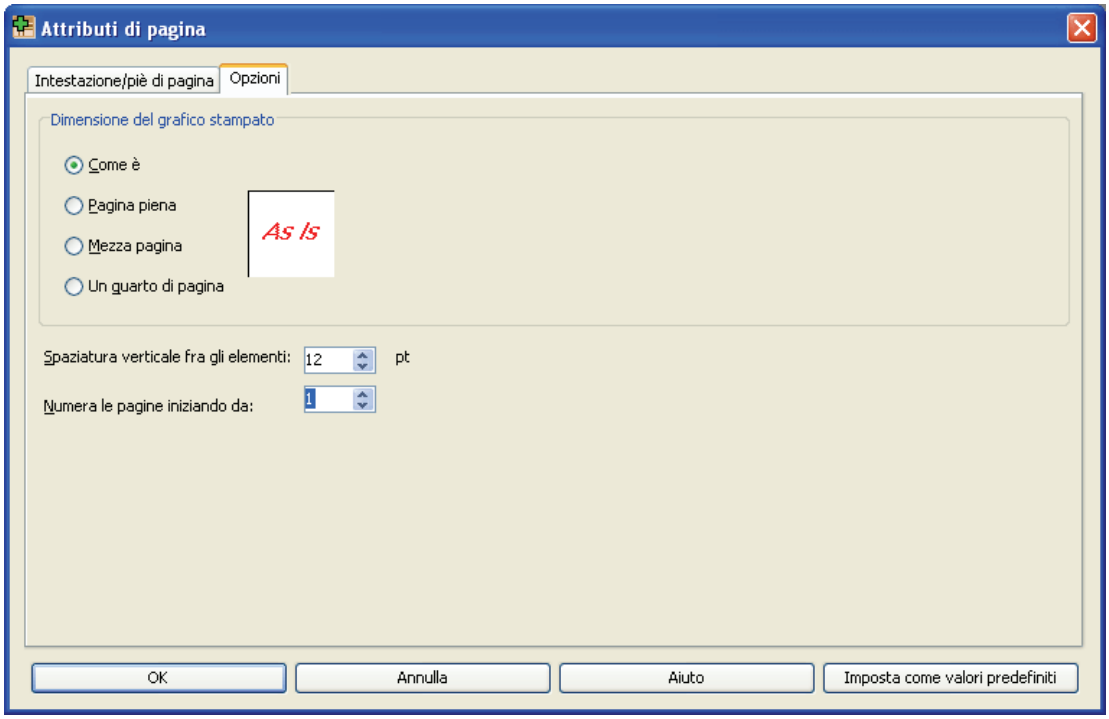
- Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).
- Dai menu, scegliere:
  File
  Attributi pagina...
- Fare clic sulla scheda Intestazione/Piè di pagina.
- Specificare l'intestazione e/o il piè di pagina che si desidera inserire in ogni pagina.

Attributi di pagina: Opzioni

In questa finestra di dialogo è possibile impostare le dimensioni del grafico stampato, lo spazio tra gli elementi dell'output stampato e la numerazione delle pagine.

- **Dimensione del grafico stampato.** Consente di impostare le dimensioni del grafico stampato in relazione a quelle definite per la pagina. Le dimensioni del grafico stampato non influiscono sul rapporto larghezza/altezza del grafico stesso. Le dimensioni totali di un grafico stampato sono limitate sia dall'altezza che dalla larghezza. Quando i bordi esterni del grafico raggiungono il bordo destro e sinistro della pagina, non è possibile aumentare ulteriormente le dimensioni del grafico per occupare in altezza la rimanente parte della pagina.
- **Spaziatura verticale fra gli elementi.** Consente di impostare lo spazio verticale fra gli elementi stampati. Ogni tabella pivot, grafico e oggetto di testo rappresenta un elemento distinto. Questa impostazione non influisce sulla visualizzazione degli elementi nel Viewer.
- **Numera le pagine iniziando da.** Le pagine vengono numerate in sequenza, iniziando dal numero indicato.
- **Imposta come valori predefiniti.** Questa opzione utilizza le impostazioni specificate qui come predefinite per i nuovi documenti Viewer. (Nota: in questo modo le impostazioni correnti sia nella scheda Intestazione/piè di pagina sia nella scheda Opzioni vengono considerate impostazioni predefinite.)

Figura 10-13
Finestra di dialogo Attributi di pagina, scheda Opzioni



Per modificare le dimensioni del grafico stampato, la numerazione delle pagine e lo spazio tra gli elementi stampati

- Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).
- Dai menu, scegliere:

  File
  Attributi pagina...
- Fare clic sulla scheda Opzioni.
- Modificare le impostazioni e quindi fare clic su OK.

Salvataggio dell'output

È possibile salvare il contenuto del Viewer come documento. Il documento salvato include entrambi i riquadri della finestra del Viewer, ovvero la vista riassuntiva e il contenuto.

Per salvare un documento Viewer

- Dai menu del Viewer, scegliere:

  File
  Salva
- Digitare il nome del documento, quindi fare clic su Salva.

Per salvare l'output in formati esterni (ad esempio, HTML o di testo), utilizzare il comando Esporta del menu File

Capitolo
11

Tabelle pivot

Molti risultati sono presentati in tabelle che possono essere modificate in modo interattivo. In altri termini, è possibile ridisporre le righe, le colonne e gli strati.

Gestione di una tabella pivot

Sono disponibili le seguenti opzioni per la gestione delle tabelle pivot:

- Trasposizione di righe e colonne
- Spostamento di righe e colonne
- Creazione di strati multidimensionali
- Raggruppamento e separazione di righe e colonne
- Possibilità di visualizzare e nascondere righe, colonne e altre informazioni
- Rotazione delle etichette delle righe e delle colonne
- Ricerca di definizioni di termini

Attivazione di una tabella pivot

Prima di poter gestire o modificare una tabella pivot, è necessario **attivare** la tabella stessa. Per attivare una tabella:

▶ Fare doppio clic sulla tabella.

o

▶ Fare clic con il pulsante destro del mouse sulla tabella e scegliere Modifica contenuto dal menu di scelta rapida.

▶ Dal sottomenu scegliere Nel Viewer o In una finestra separata.

- Per impostazione predefinita, facendo doppio clic per attivare la tabella, si attiva tutto tranne le tabelle molto grandi presenti nella finestra del Viewer. Per ulteriori informazioni, vedere Opzioni tabella pivot in Capitolo 48 a pag. 534.
- Se si desidera che più di una tabella pivot venga attivata contemporaneamente, è necessario attivare le tabelle in finestre separate.

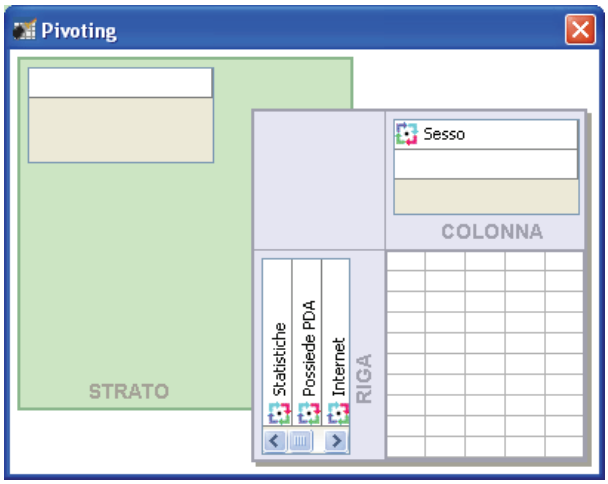
Pivoting di una tabella

▶ Attivare la tabella pivot.

- Dai menu, scegliere:

  Pivot
  Pivoting

Figura 11-1
Pivoting



Una tabella ha tre dimensioni: righe, colonne e strati. Una dimensione può contenere più elementi oppure nessun elemento. È possibile modificare l'organizzazione della tabella spostando gli elementi tra le dimensioni o all'interno di esse. Per spostare un elemento, è sufficiente trascinarlo nel punto desiderato.

Modifica dell'ordine di visualizzazione degli elementi all'interno di una dimensione

Per modificare l'ordine di visualizzazione degli elementi all'interno di una dimensione di tabella (riga, colonna o strato):

- Se il pivoting non è stato attivato, dal menu della tabella pivot, scegliere:

  Pivot
  Pivoting

- Trascinare gli elementi all'interno della dimensione nel pivoting.

Spostamento di righe e colonne all'interno di un elemento relativo alle dimensioni

- Nella tabella stessa (e non nei pivoting), fare clic sull'etichetta della riga o della colonna che si desidera spostare.

- Trascinare l'etichetta nella nuova posizione.

- Dal menu di scelta rapida selezionare Inserisci prima o Scambia.

  Nota: verificare che il comando Trascina per copiare del menu Modifica *non* sia attivato. Se il comando Trascina per copiare è attivato, deselezionarlo.

Per trasporre righe e colonne

Se si desidera invertire le righe e le colonne, ecco una semplice alternativa all'utilizzo dei pivoting:

▶ Dai menu, scegliere:

Pivot
  Trasponi righe e colonne

Questa operazione equivale a trascinare tutti gli elementi relativi alle righe nella dimensione di colonna e tutti gli elementi relativi alle colonne nella dimensione di riga.

Raggruppamento di righe o colonne

▶ Selezionare le etichette relative alle righe o alle colonne che si desidera raggruppare. Fare clic e trascinare oppure fare clic tenendo premuto Maiusc per selezionare più etichette.

▶ Dai menu, scegliere:

Modifica
  Raggruppa

Verrà inserita automaticamente un'etichetta di gruppo. Fare doppio clic sull'etichetta di gruppo per modificarne il testo.

Figura 11-2
Gruppi ed etichette di righe e colonne

		Femmina	Maschio	Totale
	Impiegato	206	157	363
	Funzionario		27	27
	Dirigente	10	74	84

Nota: per aggiungere righe o colonne a un gruppo esistente, è innanzitutto necessario separare gli elementi costitutivi del gruppo. Quindi, è possibile creare un nuovo gruppo che includa gli elementi aggiuntivi.

Separazione di righe o colonne

▶ Fare clic su qualunque punto all'interno dell'etichetta del gruppo per le righe e colonne da rimuovere dal gruppo.

▶ Dai menu, scegliere:

Modifica
  Annulla raggruppamento

Annullando il raggruppamento verrà eliminata automaticamente l'etichetta di gruppo.

Rotazione delle etichette di riga o colonna

È possibile ruotare le etichette tra la visualizzazione orizzontale e verticale per le etichette delle colonne interne e le etichette delle righe esterne presenti in una tabella.

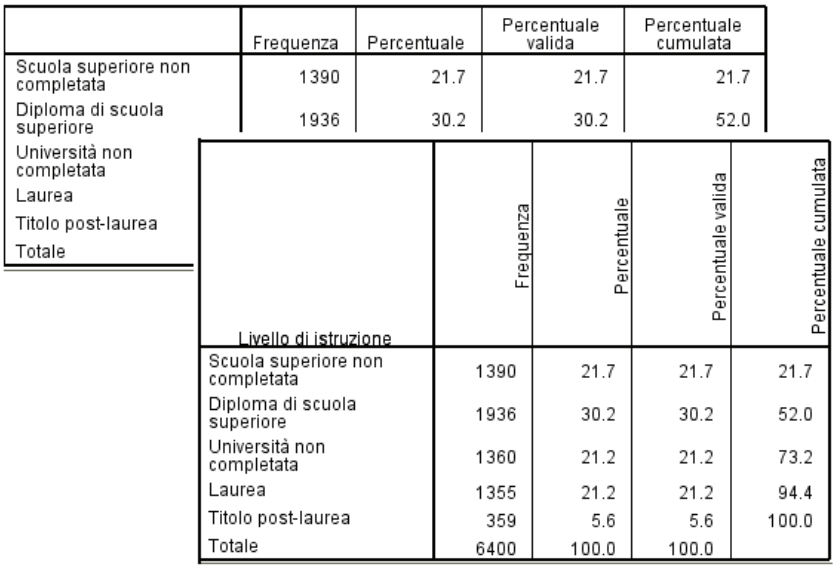
▶ Dai menu, scegliere:

Formato
  Ruota etichette colonne interne

o

Formato
  Ruota etichette righe esterne

Figura 11-3
Etichette di colonna ruotate



	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Scuola superiore non completata	1390	21.7	21.7	21.7
Diploma di scuola superiore	1936	30.2	30.2	52.0
Università non completata				
Laurea				
Titolo post-laurea				
Totale				

Livello di istruzione	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Scuola superiore non completata	1390	21.7	21.7	21.7
Diploma di scuola superiore	1936	30.2	30.2	52.0
Università non completata	1360	21.2	21.2	73.2
Laurea	1355	21.2	21.2	94.4
Titolo post-laurea	359	5.6	5.6	100.0
Totale	6400	100.0	100.0	

È possibile ruotare solo le etichette delle colonne interne e le etichette delle righe esterne.

Utilizzo degli strati

È possibile visualizzare una tabella bidimensionale distinta per ciascuna categoria o combinazione di categorie. La tabella può essere considerata come formata da diversi strati, di cui solo il primo è visibile.

Creazione e visualizzazione di strati

Per creare gli strati:

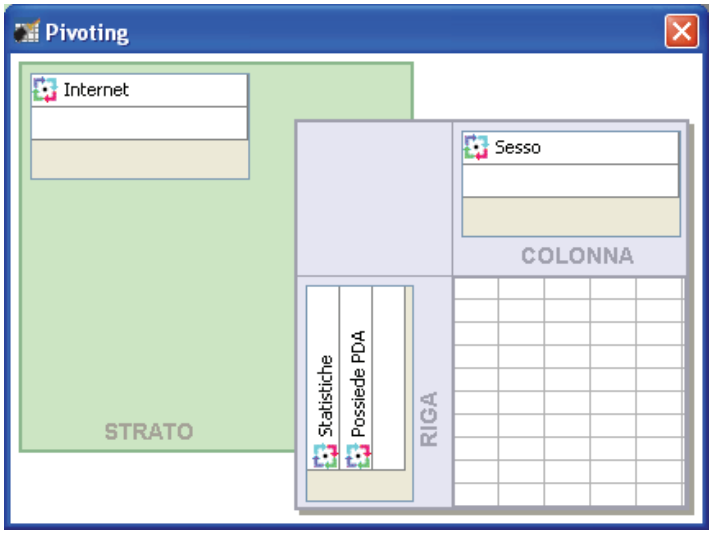
▶ Attivare la tabella pivot.

▶ Se il pivoting non è stato attivato, dal menu della tabella pivot, scegliere:

Pivot
  Pivoting

▶ Trascinare un elemento dalla dimensione di riga o di colonna nella dimensione di strato.

Figura 11-4
Spostamento di categorie negli strati



Spostando gli elementi nella dimensione di strato viene creata una tabella multidimensionale, ma viene visualizzata solo una singola "parte" bidimensionale. La tabella visibile corrisponde allo strato superiore. Ad esempio, se nella dimensione di strato è presente una variabile categoriale sì/no, la tabella multidimensionale contiene due strati: uno per la categoria *sì* e uno per la categoria *no*.

Figura 11-5
Categorie in strati distinti

Internet=Sì

		Statistiche	Sesso		
			Femmina	Maschio	Totale
Possiede PDA	No	Conteggio	513	512	1025
		Conteggio atteso	512.5	512.5	1025.0

Internet=No

		Statistiche	Sesso		
			Femmina	Maschio	Totale
Possiede PDA	No	Conteggio	1897	1962	3859
		Conteggio atteso	1903.4	1955.6	3859.0
		% entro Possiede PDA	49.2%	50.8%	100.0%
	Sì	Conteggio	327	323	650
		Conteggio atteso	320.6	329.4	650.0
		% entro Possiede PDA	50.3%	49.7%	100.0%
	Totale	Conteggio	2224	2285	4509
		Conteggio atteso	2224.0	2285.0	4509.0
		% entro Possiede PDA	49.3%	50.7%	100.0%

Modifica dello strato visualizzato

- Selezionare una categoria dall'elenco a discesa relativo agli strati (nella tabella pivot stessa e non nel pivoting).

Figura 11-6
Selezione di strati dagli elenchi a discesa

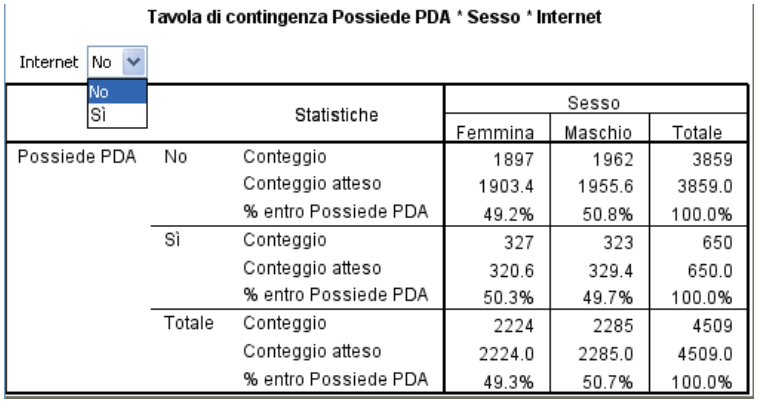


Tavola di contingenza Possiede PDA * Sesso * Internet

Internet: No (No / Sì)

		Statistiche	Sesso		
			Femmina	Maschio	Totale
Possiede PDA	No	Conteggio	1897	1962	3859
		Conteggio atteso	1903.4	1955.6	3859.0
		% entro Possiede PDA	49.2%	50.8%	100.0%
	Sì	Conteggio	327	323	650
		Conteggio atteso	320.6	329.4	650.0
		% entro Possiede PDA	50.3%	49.7%	100.0%
	Totale	Conteggio	2224	2285	4509
		Conteggio atteso	2224.0	2285.0	4509.0
		% entro Possiede PDA	49.3%	50.7%	100.0%

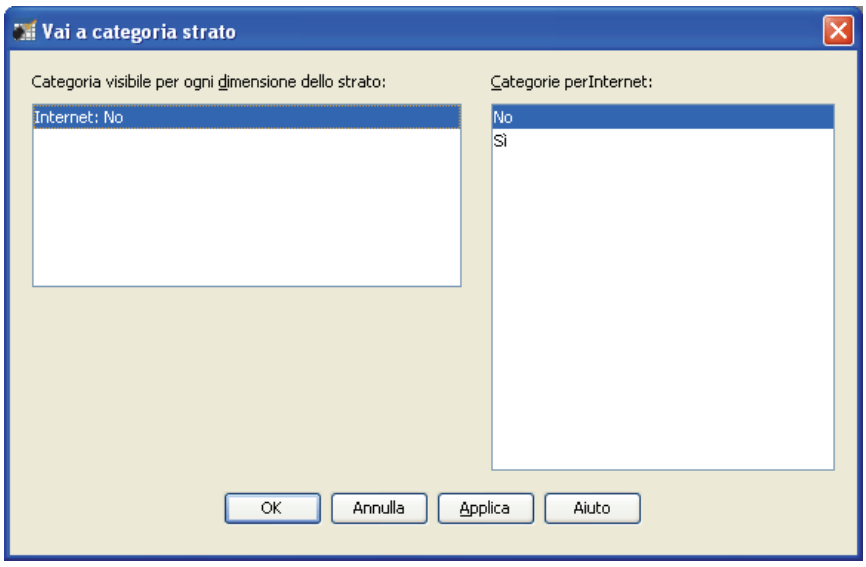
Vai a categoria strato

L'opzione Vai allo strato consente di modificare gli strati di una tabella pivot. Questa finestra di dialogo risulta particolarmente utile se è presente un grande numero di strati o se lo strato selezionato include numerose categorie.

▶ Dai menu, scegliere:

Pivot
Vai allo strato...

Figura 11-7
Finestra di dialogo Vai allo strato



▶ Nell'elenco Categoria visibile, selezionare la dimensione di uno strato. Nell'elenco Categorie verranno visualizzate tutte le categorie per la dimensione selezionata.

▶ Selezionare la categoria desiderata nell'elenco Categorie, quindi fare clic su OK o su Applica.

Visualizzare o nascondere gli elementi

È possibile nascondere molti tipi di cella, compresi:

- Etichette delle dimensioni.
- Categorie, incluse le celle dell'etichetta e dei dati di una riga o colonna.
- Etichette delle categorie, senza nascondere le celle dei dati.
- Annotazioni, titoli e didascalie.

Nascondere le righe e le colonne di una tabella

▶ Fare clic con il pulsante destro del mouse sull'etichetta della categoria relativa alla riga o alla colonna che si desidera nascondere.

▶ Dal menu di scelta rapida, scegliere:

Seleziona
  Celle dati e celle etichette

▶ Fare nuovamente clic con il pulsante destro del mouse sull'etichetta della categoria e scegliere Nascondi categoria dal menu di scelta rapida.

o

▶ Dal menu Visualizza scegliere Nascondi.

Visualizzazione di righe e colonne nascoste in una tabella

▶ Fare clic con il pulsante destro del mouse su un'altra etichetta di riga o di colonna nella stessa dimensione della riga o della colonna nascosta.

▶ Dal menu di scelta rapida, scegliere:

Seleziona
  Celle dati e celle etichette

▶ Dai menu, scegliere:

Visualizza
  Mostra tutte le categorie in nome dimensione

o

▶ Per visualizzare tutte le righe e colonne nascoste in una tabella pivot attivata, dai menu scegliere:

Visualizza
  Mostra tutto

Verranno visualizzate tutte le righe e colonne nascoste della tabella. Se è selezionata l'opzione Nascondi righe e colonne vuote nella finestra di dialogo Proprietà tabella per questa tabella, le righe o le colonne completamente vuote rimarranno nascoste.

Nascondere e visualizzare le etichette delle dimensioni

▶ Selezionare l'etichetta di dimensione o qualsiasi etichetta di categoria all'interno della dimensione.

- Dal menu Visualizza o dal menu di scelta rapida scegliere Nascondi etichetta dimensione oppure Mostra etichetta dimensione.

Nascondere e visualizzare i titoli delle tabelle

Per nascondere un titolo:

- Selezionare il titolo.
- Dal menu Visualizza scegliere Nascondi.

Per mostrare i titoli nascosti:

- Dal menu Visualizza scegliere Mostra tutto.

Modelli di tabelle

Un TableLook è un insieme di proprietà che definiscono l'aspetto di una tabella. È possibile selezionare un TableLook precedentemente definito oppure crearne uno personalizzato.

- Prima o dopo aver applicato un TableLook, è possibile modificare i formati di singole celle o di gruppi di celle utilizzando le proprietà delle celle. I formati delle celle modificati verranno mantenuti anche se si applica un nuovo TableLook. Per ulteriori informazioni, vedere Proprietà cella a pag. 256.
- È inoltre possibile ripristinare in tutte le celle i formati definiti dal TableLook corrente. In questo modo verranno ripristinate tutte le celle modificate. Se nell'elenco di file TableLook è selezionata l'opzione Com'è, le celle modificate verranno reimpostate in base alle proprietà correnti della tabella.

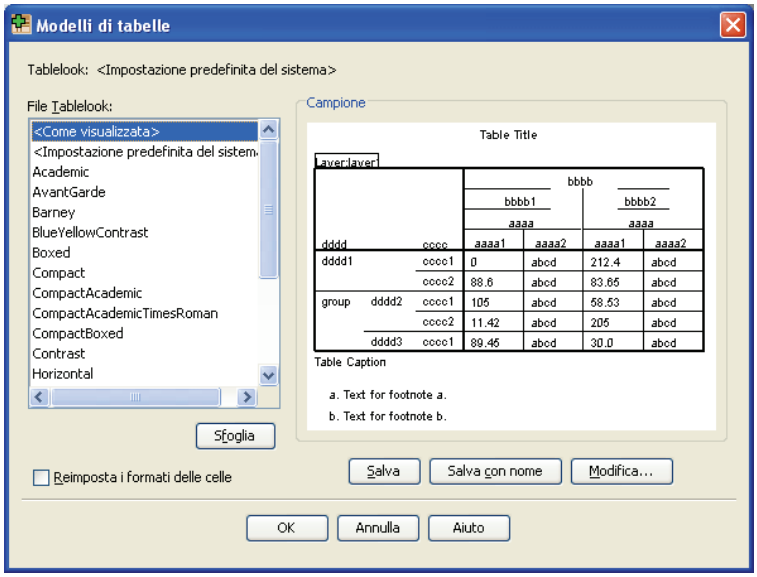
Nota: i tablelook creati nelle versioni precedenti di SPSS Statistics non possono essere utilizzati nella versione 16.0 o nelle versioni successive.

Per applicare o salvare un TableLook

- Attivare una tabella pivot.
- Dai menu, scegliere:

Formato
  Tablelook...

Figura 11-8
Finestra di dialogo TableLook



- Selezionare un TableLook dall'elenco dei file. Per selezionare un file in un'altra directory, fare clic su Sfoglia.
- Fare clic su OK per applicare il TableLook alla tabella pivot selezionata.

Per modificare o creare un TableLook

- Nella finestra di dialogo TableLooks selezionare un TableLook dall'elenco dei file.
- Fare clic su Modifica.
- Modificare le proprietà della tabella in base agli attributi desiderati e quindi fare clic su OK.
- Fare clic su Salva per salvare il TableLook modificato oppure su Salva con nome per salvarlo come nuovo TableLook.

  La modifica di un TableLook viene applicata solo alla tabella pivot selezionata. Un TableLook modificato verrà applicato alle tabelle per le quali è impostato solo se le tabelle vengono selezionate e vi viene applicato nuovamente il TableLook.

Proprietà tabella

In Proprietà tabella è possibile impostare le proprietà generali di una tabella, impostare gli stili delle celle per diverse parti della tabella e salvare un insieme di proprietà come TableLook. È possibile:

- Impostare proprietà generali, ad esempio nascondere le righe o le colonne vuote e definire le proprietà di stampa.
- Impostare il formato e la posizione dei richiami alle note.

- Determinare formati specifici per le celle nell'area dei dati, per le etichette di riga e di colonna e per altre aree della tabella.
- Impostare lo spessore e il colore delle linee che formano i bordi di ciascuna area della tabella.

Per modificare le proprietà delle tabelle pivot

- ▶ Attivare la tabella pivot.
- ▶ Dai menu, scegliere:
  Formato
  Proprietà tabella...
- ▶ Selezionare una scheda (Generale, Annotazioni, Formati di cella, Bordi o Stampa).
- ▶ Selezionare le opzioni desiderate.
- ▶ Fare clic su OK o su Applica.

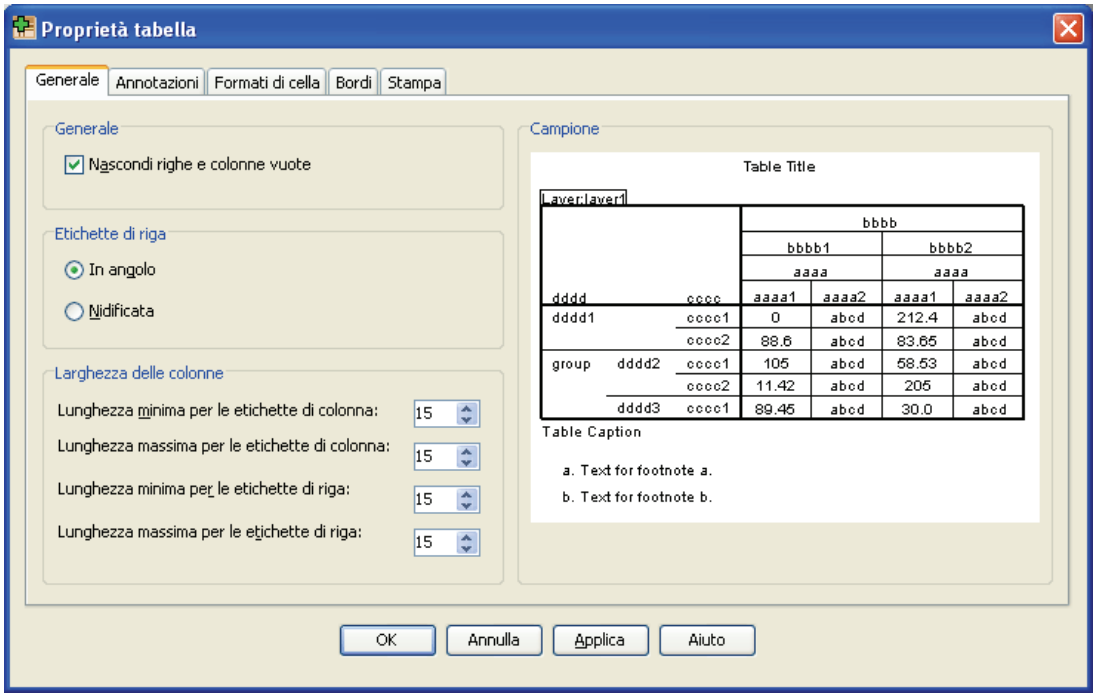
Le nuove proprietà vengono applicate alla tabella pivot selezionata. Per applicare le nuove proprietà della tabella a un TableLook anziché alla tabella selezionata, modificare il TableLook (menu Formato, TableLook).

Proprietà tabella: Generale

Diverse proprietà vengono applicate all'intera tabella. È possibile:

- Visualizzare o nascondere le righe e le colonne vuote. Una riga o una colonna vuota contiene celle di dati vuote.
- Controllare la posizione delle etichette riga, che possono essere nell'angolo in alto a sinistra o nidificate.
- Impostare la larghezza minima e massima della colonna (espressa in punti).

Figura 11-9
Finestra di dialogo Proprietà tabella, scheda Generale



Per modificare le proprietà generali di una tabella:

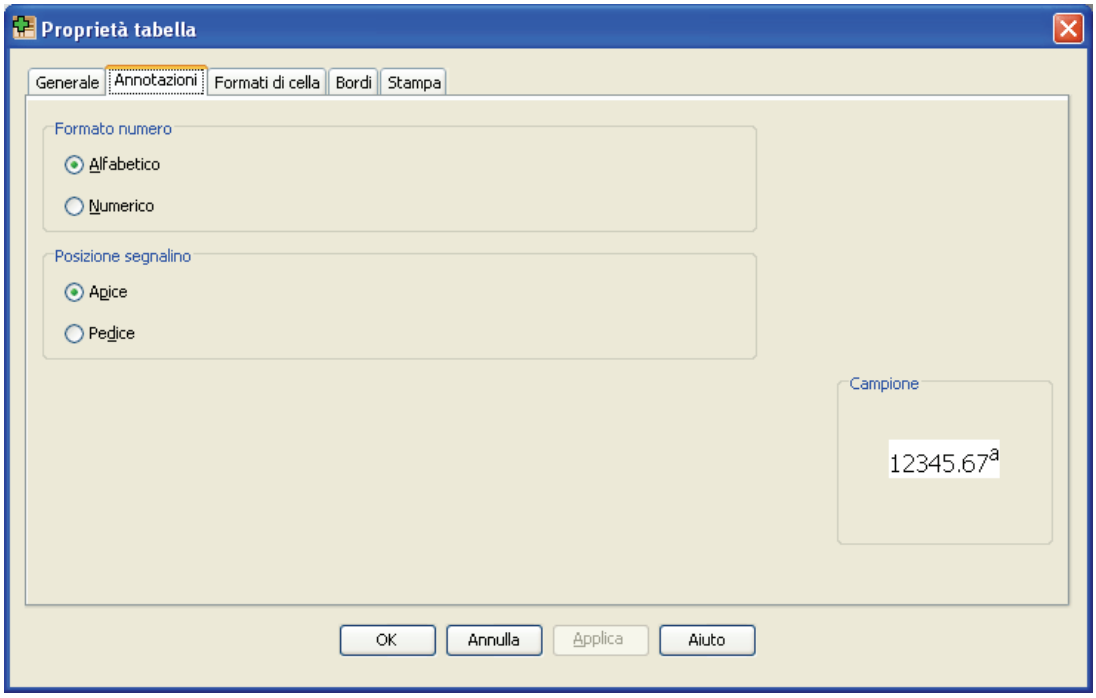
- Fare clic sulla scheda Generale.
- Selezionare le opzioni desiderate.
- Fare clic su OK o su Applica.

Proprietà tabella: Didascalie

Le proprietà dei rimandi alle note includono lo stile e la posizione in relazione al testo.

- Lo stile dei rimandi alle note può essere rappresentato da numeri (1, 2, 3...) o da lettere (a, b, c...).
- I rimandi alle note possono essere associati al testo in formato apice o pedice.

Figura 11-10
Finestra di dialogo Proprietà tabella, scheda Annotazioni



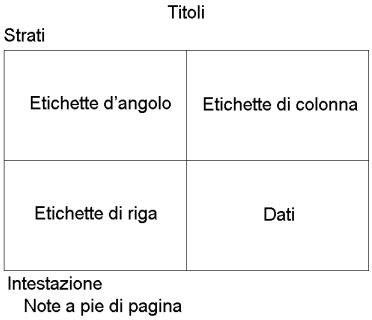
Per modificare le proprietà dei rimandi alle note:

- Selezionare la scheda Annotazioni.
- Selezionare un formato per i rimandi alle note.
- Selezionare la posizione dei rimandi.
- Fare clic su OK o su Applica.

Proprietà tabella: Formati di cella

Per quanto riguarda la formattazione, la tabella è suddivisa in più aree: Titolo, Strati, Etichette angoli, Etichette righe, Etichette colonne, Dati, Didascalia e Annotazioni. In ogni area della tabella è possibile modificare i formati di cella corrispondenti. I formati di cella includono caratteristiche del testo (ad esempio carattere, dimensione, colore, stile), allineamento orizzontale e verticale, colori di sfondo e margini interni della cella.

Figura 11-11
Aree di una tabella

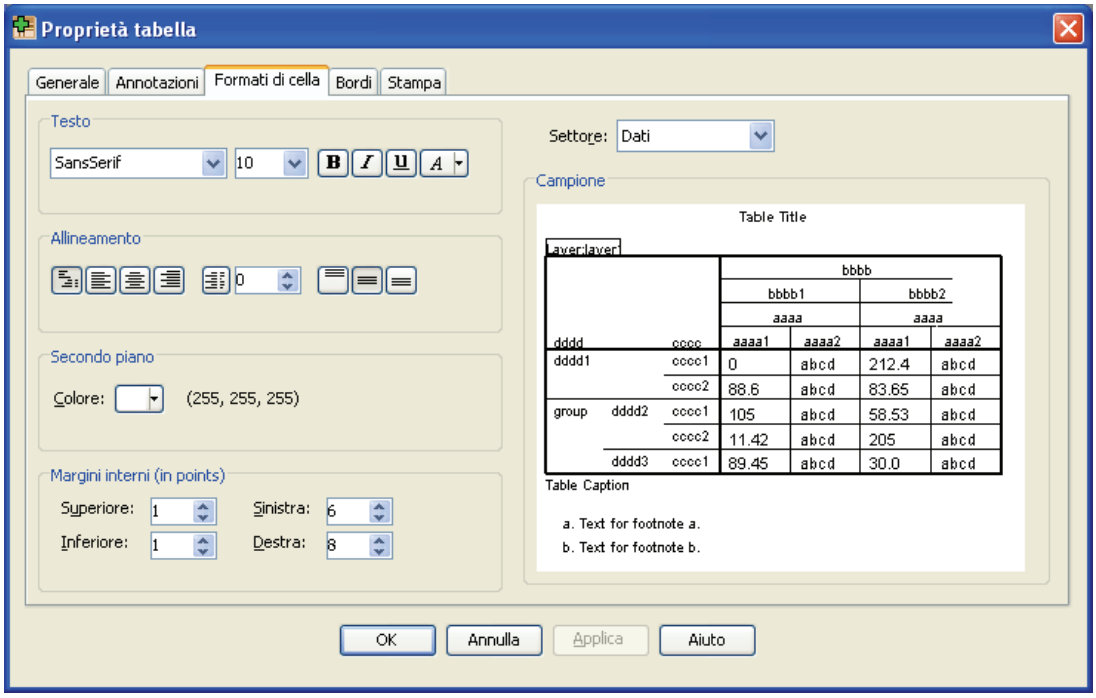


I formati di cella vengono applicati alle aree (categorie di informazioni). I formati non sono caratteristici di singole celle. Questa distinzione è importante ai fini del pivoting di una tabella.

Ad esempio,

- Se si specifica un carattere in grassetto come formato di cella per le etichette delle colonne, le etichette delle colonne appariranno in grassetto indipendentemente dal tipo di informazioni visualizzate al momento nella dimensione di colonna. Se si sposta un elemento dalla dimensione di colonna ad un'altra dimensione, tale elemento non conserverà la caratteristica relativa al grassetto delle etichette della colonna.
- Se alle etichette delle colonne viene applicato il grassetto semplicemente evidenziando le celle di una tabella pivot attivata e facendo clic sul pulsante Grassetto sulla barra degli strumenti, il contenuto delle celle rimarrà in grassetto indipendentemente dalla dimensione in cui viene spostato, mentre nelle etichette delle colonne la caratteristica grassetto non verrà applicata ad altri elementi spostati nella dimensione colonna.

Figura 11-12
Finestra di dialogo Proprietà tabella, scheda Formato celle



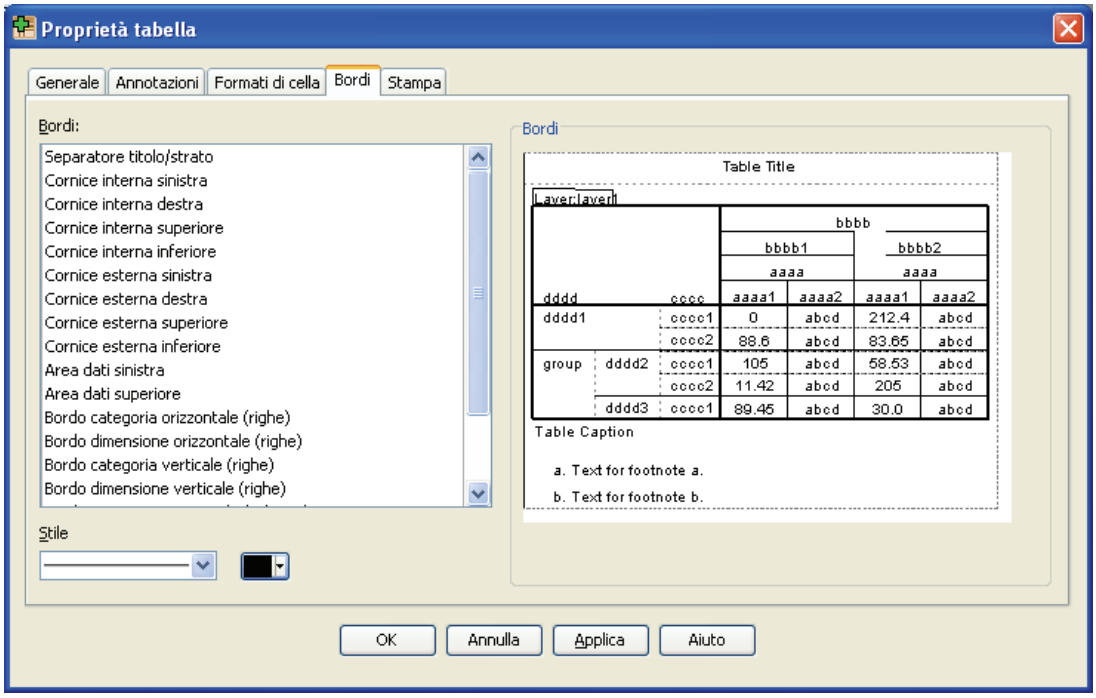
Per modificare i formati di cella:

- Selezionare la scheda Formati di cella.
- Selezionare un'area dall'elenco a discesa oppure fare clic su un area del campione.
- Selezionare le caratteristiche dell'area. Le selezioni verranno applicate al campione.
- Fare clic su OK o su Applica.

Proprietà tabella: Bordi

È possibile selezionare uno stile di linea e un colore da applicare nella posizione di ciascun bordo della tabella. Se come stile si seleziona Nessuno, nella posizione selezionata non verrà inserita alcuna linea.

Figura 11-13
Finestra di dialogo Proprietà tabella, scheda Bordi



Per modificare i bordi di una tabella:

- Fare clic sulla scheda Bordi.
- Selezionare la posizione del bordo facendo clic sul suo nome nell'elenco oppure facendo clic su una linea nell'area Esempio.
- Selezionare uno stile di linea oppure Nessuno.
- Selezionare un colore.
- Fare clic su OK o su Applica.

Proprietà tabella: Stampa

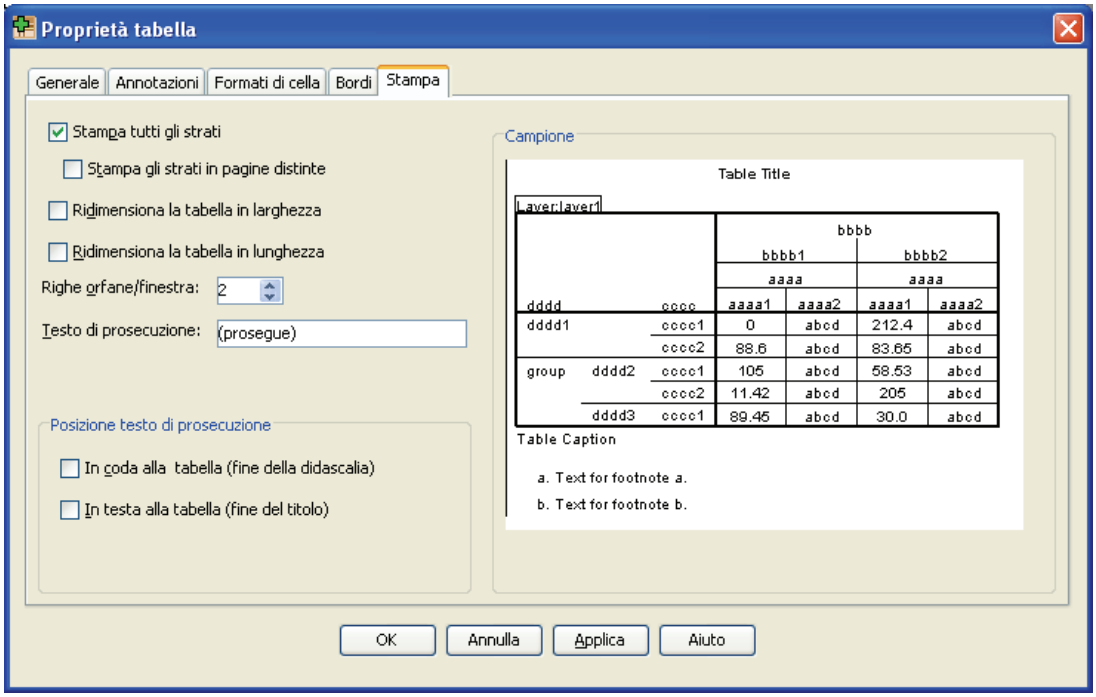
Per le tabelle pivot stampate è possibile definire quanto segue:

- Stampare tutti gli strati o solo lo strato superiore di una tabella e stampare ciascuno strato in una pagina distinta.
- Ridurre una tabella orizzontalmente o verticalmente in base alle dimensioni della pagina stampata.
- Controllare le righe vedove/orfane, ovvero il numero minimo di righe e colonne che verranno incluse in una sezione stampata di una tabella, se questa ha lunghezza o larghezza eccessive per le dimensioni definite per la pagina.

Nota: se una tabella è troppo lunga per poter essere inclusa nella pagina corrente solo perché è preceduta da altro output, verrà stampata automaticamente in una nuova pagina, indipendentemente dall'impostazione delle righe vedove/orfane.

- Includere testo di prosecuzione per le tabelle che non rientrano in una sola pagina. È possibile visualizzare il testo di prosecuzione in testa o in coda a ciascuna pagina. Se nessuna delle due opzioni è selezionata, il testo di prosecuzione non sarà visualizzato.

Figura 11-14
Finestra di dialogo Proprietà tabella, scheda Stampa



Per controllare le proprietà di stampa delle tabelle pivot:

- ▶ Fare clic sulla scheda Stampa.
- ▶ Selezionare le opzioni di stampa desiderate.
- ▶ Fare clic su OK o su Applica.

Proprietà cella

Le opzioni di Proprietà cella vengono applicate a una singola cella selezionata. È possibile modificare il carattere, il formato dei valori, l'allineamento, i margini e i colori. Le opzioni di Proprietà cella hanno la priorità sulle opzioni di Proprietà tabella. Pertanto, se si modificano le proprietà della tabella, non vengono modificate le proprietà delle celle applicate singolarmente.

Per modificare le proprietà delle celle:

- ▶ Attivare una tabella e selezionare le celle al suo interno.
- ▶ Dal menu Formato o dal menu di scelta rapida selezionare Proprietà cella.

Carattere e sfondo

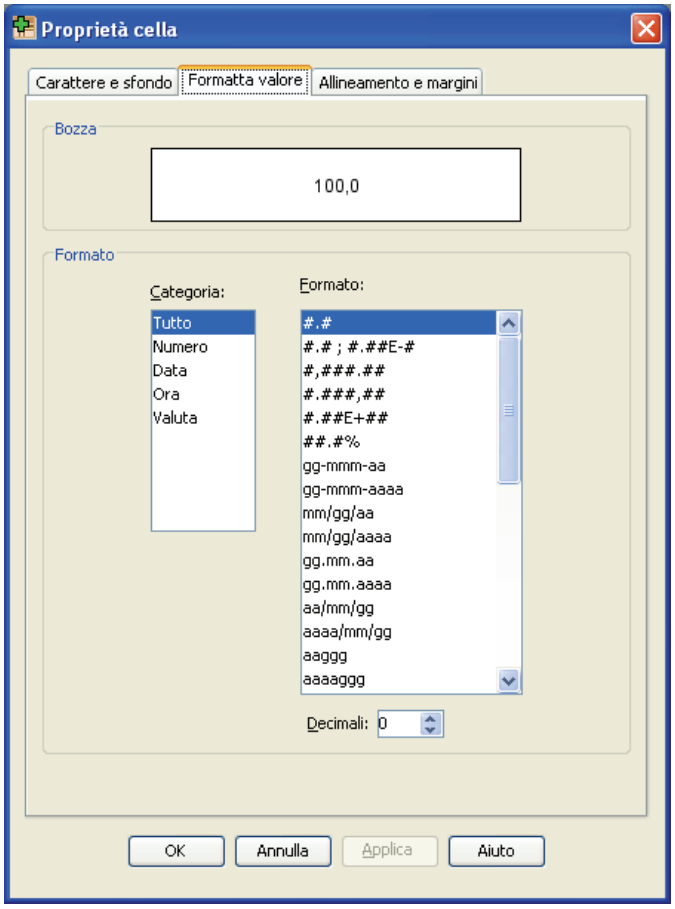
Nella scheda Carattere e sfondo è possibile controllare lo stile del carattere, il colore e il colore di sfondo per le celle selezionate nella tabella.

Figura 11-15
Finestra di dialogo Proprietà cella, scheda Carattere e sfondo

Formatta valore

Nella scheda Formatta valore è possibile controllare il formato dei valori delle celle selezionate. È possibile selezionare il formato dei numeri, della data, dell'orario e della valuta e impostare il numero di cifre decimali da visualizzare.

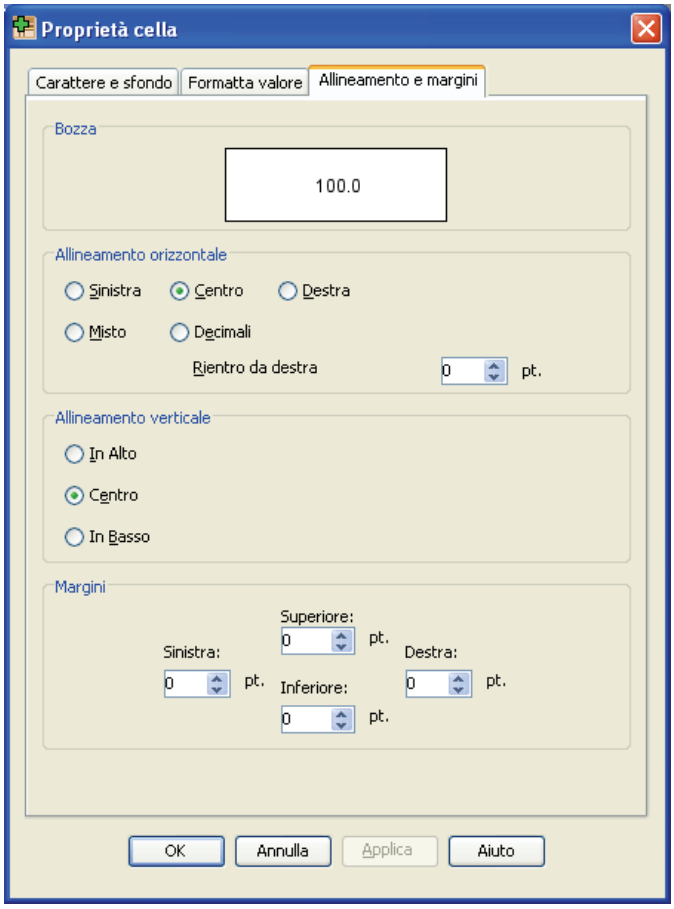
Figura 11-16
Finestra di dialogo Proprietà cella, scheda Formatta valore



Allineamento e margini

Nella scheda Allineamento e margini è possibile controllare l'allineamento orizzontale e verticale dei valori e i margini superiore, inferiore, sinistro e destro per le celle selezionate. L'allineamento orizzontale Misto consente di allineare il contenuto di ogni cella in base al tipo. Ad esempio, le date sono allineate a destra e i valori di testo sono allineate a sinistra.

Figura 11-17
Finestra di dialogo Proprietà cella, scheda Allineamento e margini



Annotazioni e didascalie

È possibile aggiungere annotazioni e didascalie a una tabella. È inoltre possibile nascondere le annotazioni o le didascalie, modificare i rimandi alle annotazioni e rinumerare le annotazioni.

Alcuni attributi delle annotazioni sono controllati dalle proprietà delle tabelle. Per ulteriori informazioni, vedere Proprietà tabella: Didascalie a pag. 251.

Aggiunta di annotazioni e didascalie

Per aggiungere una didascalia a una tabella:

- Dal menu Inserisci scegliere Didascalia.

È possibile aggiungere un'annotazione a qualsiasi elemento di una tabella. Per aggiungere un'annotazione:

- Fare clic su un titolo, su una cella o su una didascalia all'interno di una tabella pivot attivata.

- Dal menu Inserisci scegliere Annotazioni.

Per visualizzare o nascondere una didascalia

Per nascondere una didascalia:

- Selezionare la didascalia.
- Dal menu Visualizza scegliere Nascondi.

Per mostrare le didascalie nascoste:

- Dal menu Visualizza scegliere Mostra tutto.

Per nascondere o visualizzare un'annotazione in una tabella

Per nascondere un'annotazione:

- Selezionare l'annotazione.
- Dal menu Visualizza scegliere Nascondi oppure dal menu di scelta rapida scegliereNascondi annotazioni.

Per mostrare le annotazioni nascoste:

- Dal menu Visualizza scegliere Mostra tutte le annotazioni.

Rimando a nota

L'opzione Rimando a nota consente di modificare i caratteri utilizzati nei rimandi alle note.

Figura 11-18
Finestra di dialogo Rimando a nota

Per modificare i rimandi alle note:

- Selezionare un'annotazione.
- Dal menu Formato scegliere Rimando a nota.
- Digitare uno o due caratteri.

Per rinumerare le annotazioni

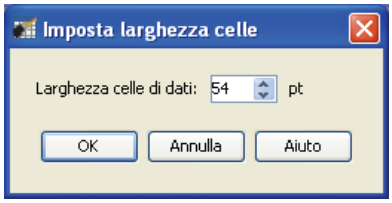
Dopo aver effettuato il pivoting di una tabella modificando righe, colonne e strati, le annotazioni possono risultare ordinate diversamente. Per rinumerare le annotazioni:

- Dal menu Formato scegliere Rinumera annotazioni.

Larghezza delle celle

L'opzione Imposta larghezza celle consente di impostare la stessa larghezza per tutte le celle di dati.

Figura 11-19
Finestra di dialogo Imposta larghezza celle



Per impostare la larghezza di tutte le celle di dati:

- Dai menu, scegliere:

  Formato
  Imposta larghezza celle...

- Specificare il valore della larghezza delle celle.

Modifica della larghezza di colonna

- Fare clic e trascinare il bordo di colonna.

Visualizzazione dei bordi nascosti in una tabella pivot

Per le tabelle in cui non sono visibili molti bordi, è possibile visualizzare i bordi nascosti. In questo modo è possibile eseguire più rapidamente alcune operazioni, ad esempio la modifica della larghezza delle colonne.

- Dal menu Visualizza scegliere Griglia.

Figura 11-20
Griglie visualizzate per i bordi nascosti

Internet				Sesso		
				Femmina	Maschio	Totale
No	Possiede PDA	No	Conteggio	1897	1962	3859
			Conteggio atteso	1903.4	1955.6	3859.0
			% entro Possiede PDA	49.2%	50.8%	100.0%
		Sì	Conteggio	327	323	650
Sì	Possiede PDA					

Internet				Sesso		
				Femmina	Maschio	Totale
No	Possiede PDA	No	Conteggio	1897.0%	1962.0%	3859.0%
			Conteggio atteso	1903.4	1955.6	3859.0
			% entro Possiede PDA	49.2%	50.8%	100.0%
		Sì	Conteggio	327.0%	323.0%	650.0%
			Conteggio atteso	320.6	329.4	650.0
			% entro Possiede PDA	50.3%	49.7%	100.0%
		Totale	Conteggio	2224.0%	2285.0%	4509.0%
			Conteggio atteso	2224.0	2285.0	4509.0
			% entro Possiede PDA	49.3%	50.7%	100.0%
Yes	Possiede PDA	No	Conteggio	513.0%	512.0%	1025.0%
			Conteggio atteso	512.5	512.5	1025.0
			% entro Possiede PDA	50.0%	50.0%	100.0%
		Sì	Conteggio	305.0%	306.0%	611.0%
			Conteggio atteso	305.5	305.5	611.0
			% entro Possiede PDA	49.9%	50.1%	100.0%
		Totale	Conteggio	818.0%	818.0%	1636.0%
			Conteggio atteso	818.0	818.0	1636.0
			% entro Possiede PDA	50.0%	50.0%	100.0%

Selezione di righe e colonne in una tabella pivot

Le tabelle pivot pongono alcune limitazioni sulla selezione di intere righe e colonne e l'evidenziazione che indica la riga o la colonna selezionata può essere estesa ad aree non contigue della tabella. Per selezionare un'intera riga o colonna:

- Fare clic su un'etichetta di riga o di colonna.
- Dai menu, scegliere:

  Modifica
  Seleziona
  Celle dati e celle etichette

  o

- Fare clic con il pulsante destro del mouse sull'etichetta di categoria relativa alla riga o alla colonna.
- Dal menu di scelta rapida, scegliere:

  Seleziona
  Celle dati e celle etichette

  o

- Fare clic tenendo premuti Ctrl e Alt sull'etichetta della riga o della colonna.

Stampa delle tabelle pivot

I fattori che influenzano la stampa delle tabelle pivot sono molteplici e possono essere definiti modificando gli attributi delle tabelle pivot.

- Per le tabelle pivot multidimensionali (tabelle con strati), è possibile scegliere se stampare tutti gli strati o soltanto il primo strato (visibile). Per ulteriori informazioni, vedere Proprietà tabella: Stampa a pag. 255.
- Per le tabelle pivot lunghe o grandi, è possibile ridefinirne automaticamente le dimensioni in modo da adattarle alla pagina oppure impostare con esattezza le interruzioni delle tabelle o delle pagine. Per ulteriori informazioni, vedere Proprietà tabella: Stampa a pag. 255.
- Per le tabelle che sono troppo larghe o troppo lunghe per rientrare in un'unica pagina, è possibile controllare la posizione delle interruzioni delle tabelle fra le pagine.

Scegliere Anteprima di stampa del menu File per visualizzare le tabelle pivot prima di stamparle.

Impostazione delle interruzioni di pagina per le tabelle con larghezza o lunghezza eccessiva

Le tabelle pivot con larghezza o altezza eccessiva per essere stampate entro le dimensioni definite per la pagina verranno divise automaticamente e stampate in più sezioni. Per le tabelle di larghezza eccessiva, le diverse sezioni verranno stampate nella stessa pagina se è disponibile spazio sufficiente. È possibile:

- Impostare i punti delle righe e delle colonne in cui le tabelle vengono divise.
- Specificare le righe e le colonne che devono essere mantenute insieme quando le tabelle vengono divise.
- Ridimensionare le tabelle di grandi dimensioni in modo da adattarle alla pagina.

Per specificare le interruzioni di riga e di colonna per le tabelle pivot da stampare

▶ Fare clic sull'etichetta di colonna o di riga situata rispettivamente a sinistra o al di sopra del punto in cui si desidera inserire l'interruzione.

▶ Dai menu, scegliere:

Formato
Interrompi qui

Per specificare le righe o le colonne da mantenere unite

▶ Selezionare le etichette delle righe o delle colonne che si desidera mantenere unite. Fare clic e trascinare oppure fare clic tenendo premuto Maiusc per selezionare più etichette di riga o di colonna.

▶ Dai menu, scegliere:

Formato
Mantieni unito

Creazione di un grafico da una tabella pivot

- Fare doppio clic nella tabella pivot per attivarla.
- Selezionare le righe, le colonne o le celle che si desidera visualizzare nel grafico.
- Fare clic con il pulsante destro del mouse in un punto qualsiasi all'interno dell'area selezionata.
- Scegliere Crea grafico dal menu di scelta rapida, quindi selezionare un tipo di grafico.

Capitolo
12

Modelli

Alcuni risultati vengono presentati come modelli, visualizzati in Output Viewer come tipo di visualizzazione speciale. La visualizzazione mostrata in Output Viewer non è l'unica vista disponibile del modello. Un unico modello contiene molte viste diverse. È possibile attivare il modello nel Viewer modelli e interagire direttamente con il modello per visualizzare le viste dei modelli disponibili. È anche possibile scegliere di stampare ed esportare tutte le viste del modello.

Interazione con un modello

Per interagire con un modello, è necessario per prima cosa attivarlo:

- Fare doppio clic sul modello.

  o

- Fare clic con il pulsante destro del mouse e scegliere Modifica contenuto dal menu di scelta rapida.
- Scegliere In una finestra separata dal sottomenu.

Quando si attiva il modello, questo viene visualizzato nel Viewer modelli. Per ulteriori informazioni, vedere Uso del Viewer modelli a pag. 265.

Uso del Viewer modelli

Il Viewer modelli è uno strumento interattivo per visualizzare le viste dei modelli disponibili e modificarne l'aspetto. Per informazioni sulla visualizzazione del Viewer modelli, vedere Interazione con un modello a pag. 265. Il Viewer modelli è diviso in due parti:

- **Vista principale**. La vista principale si trova nella parte sinistra del Viewer modelli e mostra una visualizzazione generale (ad esempio, un grafico di rete) per il modello. La vista principale a sua volta può avere più di una vista del modello. L'elenco a discesa posto sotto la vista principale consente di scegliere tra le viste principali disponibili.
- **Vista ausiliaria**. La vista ausiliaria si trova nella parte destra del Viewer modelli e generalmente mostra una visualizzazione più dettagliata (comprese le tabelle) del modello rispetto alla visualizzazione generale della vista principale. Come la vista principale, anche la vista ausiliaria può avere più di una vista del modello. L'elenco a discesa posto sotto la vista ausiliaria consente di scegliere tra le viste principali disponibili. Questa vista può anche mostrare visualizzazioni specifiche per elementi selezionati nella vista principale. Ad esempio, a seconda del tipo di modello, potrebbe essere possibile selezionare un nodo delle variabili nella vista principale per visualizzare una tabella di tale variabile nella vista ausiliaria.

Le visualizzazioni specifiche mostrate dipendono dalla procedura di creazione del modello. Per informazioni sull'uso di modelli specifici, consultare la documentazione relativa alla procedura di creazione del modello.

Tabelle del Viewer modelli

Le tabelle visualizzate nel Viewer modelli non sono tabelle pivot e non possono essere manipolate come le tabelle pivot.

Modifica ed esplorazione delle viste dei modelli

Per impostazione predefinita, il Viewer modelli viene attivato in **modalità Esplora**, che consente di interagire con il modello trascinando elementi e selezionandoli per mostrare la visualizzazione relativa nella vista ausiliaria.

Il Viewer modelli dispone inoltre di una **modalità Modifica**, che consente di modificare gli stili delle viste dei modelli. Il Viewer modelli comprende l'Editor lavagna grafica, che serve ad apportare tali modifiche. Quando si chiude il Viewer modelli, tutte le modifiche apportate vengono salvate nel modello.

Il menu Visualizza consente di passare dalla modalità Esplora alla modalità Modifica.

Impostazione delle proprietà dei modelli

All'interno del Viewer modelli è possibile impostare proprietà specifiche per il modello. Per ulteriori informazioni, vedere Proprietà dei modelli a pag. 266.

Copia delle viste dei modelli

All'interno del Viewer modelli è anche possibile copiare singole viste dei modelli. Per ulteriori informazioni, vedere Copia delle viste dei modelli a pag. 267.

Proprietà dei modelli

Dai menu del Viewer modelli, scegliere:

File
  Proprietà

Ogni modello dispone di proprietà associate che consentono di specificare le viste che vengono stampate dall'Output Viewer. Per impostazione predefinita, viene stampata solo la vista visibile nell'Output Viewer. Si tratta sempre di una sola vista principale. È anche possibile specificare la stampa di tutte le viste dei modelli disponibili, che comprendono tutte le viste principali e ausiliarie. Si noti che è anche possibile stampare singole viste dei modelli dal Viewer modelli stesso. Per ulteriori informazioni, vedere Stampa di un modello a pag. 267.

Copia delle viste dei modelli

Dal menu Modifica del Viewer modelli è possibile copiare la vista principale o la vista ausiliaria attualmente visualizzata. Viene copiata una sola vista del modello. È possibile incollare la vista del modello nell'Output Viewer, dove viene successivamente mostrata come visualizzazione modificabile nell'Editor lavagna grafica. Incollando nell'Output Viewer è possibile visualizzare più viste dei modelli contemporaneamente. È anche possibile incollare in altre applicazioni, dove la vista può apparire come un'immagine o una tabella, a seconda dell'applicazione di destinazione.

Stampa di un modello

Stampa dal Viewer modelli

Dal Viewer modelli è possibile stampare un'unica vista del modello.

- Attivare il modello nel Viewer modelli. Per ulteriori informazioni, vedere Interazione con un modello a pag. 265.
- Dai menu, scegliere:
  Visualizza
  Modalità Modifica
- Nella tavolozza Generale della vista principale o ausiliaria (a seconda di quella che si desidera stampare), fare clic sull'icona di stampa. Se la tavolozza non è visualizzata, selezionare Tavolozza>Generale dal menu Visualizza.

Stampa dall'Output Viewer

Quando si stampa dall'Output Viewer, il numero di viste stampate per un modello specifico dipende dalle proprietà del modello. Il modello può essere impostato per stampare solo la vista visualizzata oppure tutte le viste dei modelli disponibili. Per ulteriori informazioni, vedere Proprietà dei modelli a pag. 266.

Esportazione di un modello

Per impostazione predefinita, quando si esportano modelli dall'Output Viewer, l'inclusione o l'esclusione delle viste dei modelli è controllata dalle proprietà di ogni modello. Per ulteriori informazioni sulle proprietà dei modelli, vedere Proprietà dei modelli a pag. 266. Per l'esportazione è possibile ignorare questa impostazione e includere tutte le viste dei modelli o solo la vista del modello attualmente visibile. Nella finestra di dialogo Esporta output, fare clic su Opzioni di modifica... nel gruppo Documento. Per ulteriori informazioni sull'esportazione e su questa finestra di dialogo, vedere Esporta output a pag. 223. Si noti che tutte le viste dei modelli, comprese le tabelle, vengono esportate come immagini.

Capitolo
13

Utilizzo della sintassi dei comandi

Il linguaggio a comandi consente di salvare e automatizzare molte delle operazioni più comuni. Inoltre, offre alcune funzionalità a cui non è possibile accedere tramite i menu e le finestre di dialogo.

È possibile accedere alla maggior parte dei comandi tramite i menu e le finestre di dialogo. Tuttavia, alcuni comandi e opzioni sono disponibili solo utilizzando il linguaggio dei comandi. Questo linguaggio consente inoltre di salvare le sessioni in un file di sintassi, in modo da poter ripetere l'analisi successivamente oppure eseguirla in una sessione automatizzata con l'utilità SPSS Production.

Un file di sintassi è un semplice file di testo che contiene comandi. Sebbene sia possibile aprire una finestra di sintassi e digitarvi dei comandi, risulta senz'altro più semplice lasciare al programma il compito di creare il file di sintassi utilizzando uno dei seguenti metodi:

- Incollando la sintassi dei comandi dalle finestre di dialogo
- Copiando la sintassi dal registro dell'output
- Copiando la sintassi dal file giornale

Informazioni dettagliate sui riferimenti alla sintassi dei comandi sono disponibili in due formati: nella Guida integrata e su un documento PDF a parte in *Command Syntax Reference*, anch'esso consultabile tramite il menu ?.

Per visualizzare la Guida sensibile al contesto per il comando visualizzato nella finestra della sintassi, premere il tasto F1.

Regole della sintassi dei comandi

Quando si eseguono i comandi dalla finestra della sintassi di comando durante una sessione, i comandi vengono eseguiti in modalità **interattiva**.

Le seguenti regole si applicano alle specifiche dei comandi in modalità interattiva:

- Ciascun comando deve iniziare su una nuova riga. I comandi possono iniziare in qualsiasi colonna di una riga di comando e continuare per il numero di righe necessario. L'unica eccezione è rappresentata dal comando `END DATA`, che deve iniziare nella prima colonna della prima riga dopo la fine dei dati.
- Ciascun comando deve terminare con un punto. È sempre consigliabile omettere la fine del comando con `BEGIN DATA` per essere certi che i dati in linea vengano considerati come specifica continua.
- La fine del comando deve sempre essere l'ultimo carattere del comando, spazi esclusi.
- Se il comando non termina con un punto, la linea vuota viene interpretata come fine del comando.

Nota: per garantire la compatibilità con le altre modalità di esecuzione dei comandi (compresi i file di comando eseguiti con i comandi `INSERT` o `INCLUDE` in modalità interattiva), è indispensabile che ciascuna riga della sintassi di comando non superi i 256 byte di lunghezza.

- La maggior parte dei sottocomandi vengono separati da barre (/). La barra che precede il primo sottocomando di un comando è in genere opzionale.
- I nomi delle variabili devono essere riportati per intero.
- Il testo racchiuso tra apostrofi o virgolette deve occupare una sola riga.
- Indipendentemente dalle impostazioni internazionali di Windows, il separatore dei decimali deve essere un punto (.).
- I nomi di variabili che terminano con un punto possono causare errori nei comandi creati nelle finestre di dialogo. Nelle finestre di dialogo non è possibile creare nomi di variabili con un punto finale ed in genere è necessario evitarne l'utilizzo.

Per la sintassi dei comandi non si fa distinzione tra maiuscole e minuscole e, per molti dei comandi, è possibile utilizzare delle abbreviazioni di tre o quattro lettere. Per specificare un singolo comando è possibile utilizzare un numero illimitato di righe. È possibile aggiungere uno spazio o delle interruzioni di riga in qualsiasi punto in cui sia ammesso l'utilizzo di un singolo spazio vuoto, come prima e dopo le barre, le parentesi e gli operatori aritmetici o tra i nomi di variabili. Ad esempio,

```
FREQUENCIES
  VARIABLES=JOBCAT GENDER
  /PERCENTILES=25 50 75
  /BARCHART.
```

e

```
freq var=jobcat gender /percent=25 50 75 /bar.
```

sono entrambe alternative accettabili che generano i medesimi risultati.

File INCLUDE

Per i comandi eseguiti tramite il comando `INCLUDE` si applicano le regole della sintassi della modalità batch.

Le seguenti regole si applicano alle specifiche dei comandi in modalità batch:

- Tutti i comandi nel file di comandi devono iniziare dalla colonna 1; è possibile usare i simboli più (+) o meno (–) nella prima colonna per inserire un rientro per la specifica del comando in modo da renderlo più leggibile.
- Se si utilizzano più righe per il comando, la colonna 1 di ciascuna riga successiva deve essere lasciata vuota.
- I terminatori dei comandi sono opzionali.
- Una riga deve avere una lunghezza massima di 256 byte; i caratteri aggiuntivi vengono troncati.

Se non si hanno file di comandi che già utilizzano il comando INCLUDE, è preferibile usare il comando INSERT poiché può essere utilizzato con file di comando conformi ad entrambi gli insiemi di regole. Se la sintassi viene generata incollando le selezioni nelle finestre di dialogo in una finestra di sintassi, il formato dei comandi è idoneo per qualsiasi modalità. Per ulteriori informazioni, vedere i *riferimenti sulla sintassi dei comandi*, disponibili in formato PDF nel menu ?.

Creazione di sintassi dei comandi nelle finestre di dialogo

Il metodo più semplice per creare un file di sintassi consiste nell'eseguire selezioni nelle finestre di dialogo di SPSS e quindi incollare la sintassi per le selezioni in una finestra di sintassi. Incollando la sintassi in ciascuna fase di una lunga analisi, si costruisce un file della sessione che permette di ripetere l'analisi in un secondo momento o di eseguire una sessione automatizzata con l'utilità SPSS Production.

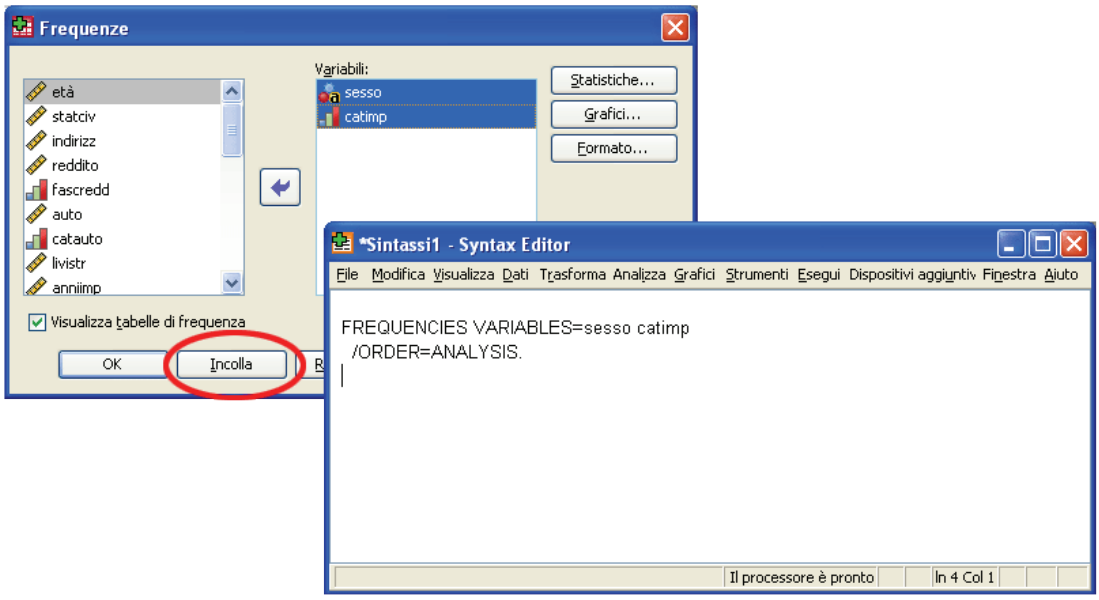
Nella finestra di sintassi, è possibile eseguire la sintassi incollata, modificarla e salvarla in un file di sintassi.

Per creare sintassi dei comandi nelle finestre di dialogo

- Aprire la finestra di dialogo ed eseguire tutte le selezioni desiderate.
- Fare clic su Incolla.

La sintassi dei comandi viene incollata nella finestra di sintassi designata. Se non è aperta alcuna finestra di sintassi, ne viene automaticamente aperta una nuova in cui viene incollata la sintassi copiata.

Figura 13-1
Sintassi incollata da una finestra di dialogo

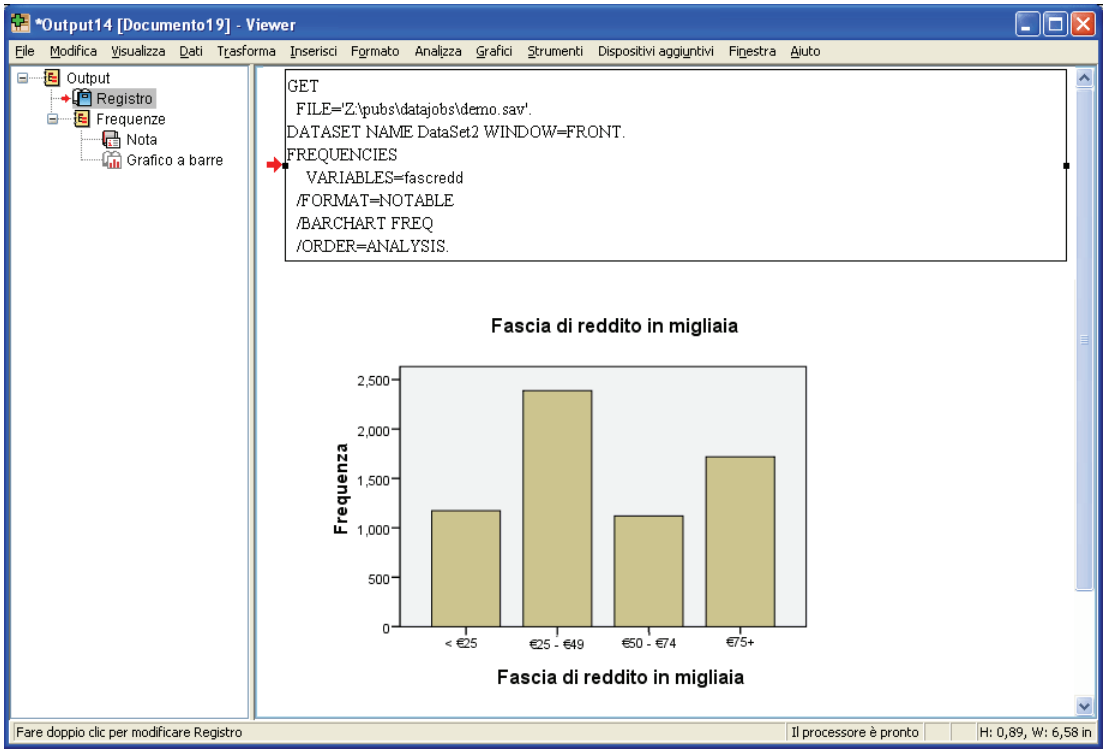


Copia della sintassi dal registro dell'output

È possibile costruire un file di sintassi copiando la sintassi dei comandi dal registro visualizzato nel Viewer. Per utilizzare questo metodo, selezionare Visualizza i comandi nel registro nelle impostazioni del Viewer (menu Modifica, Opzioni, scheda Viewer) prima di eseguire l'analisi. In tal modo, tutti i comandi appariranno nel Viewer insieme all'output dell'analisi.

Nella finestra di sintassi, è possibile eseguire la sintassi incollata, modificarla e salvarla in un file di sintassi.

Figura 13-2
Sintassi dei comandi nel registro



Per copiare la sintassi dal registro dell'output

- Prima di eseguire l'analisi, dai menu scegliere:

  Modifica
  Opzioni...

- Nella scheda Viewer, selezionare Visualizza i comandi nel registro.

  All'esecuzione delle analisi, i comandi relativi alle selezioni effettuate nella finestra di dialogo vengono inseriti nel registro.

- Aprire un file di sintassi salvato in precedenza oppure crearne uno nuovo. Per creare un nuovo file di sintassi, dai menu scegliere:

  File
  Nuovo
  Sintassi

- Nel Viewer, fare doppio clic su una voce del registro per attivarla.

- Selezionare il testo da copiare.

- Dai menu del Viewer scegliere:

  Modifica
  Copia

- In una finestra di sintassi, dai menu scegliere:

  Modifica
  Incolla

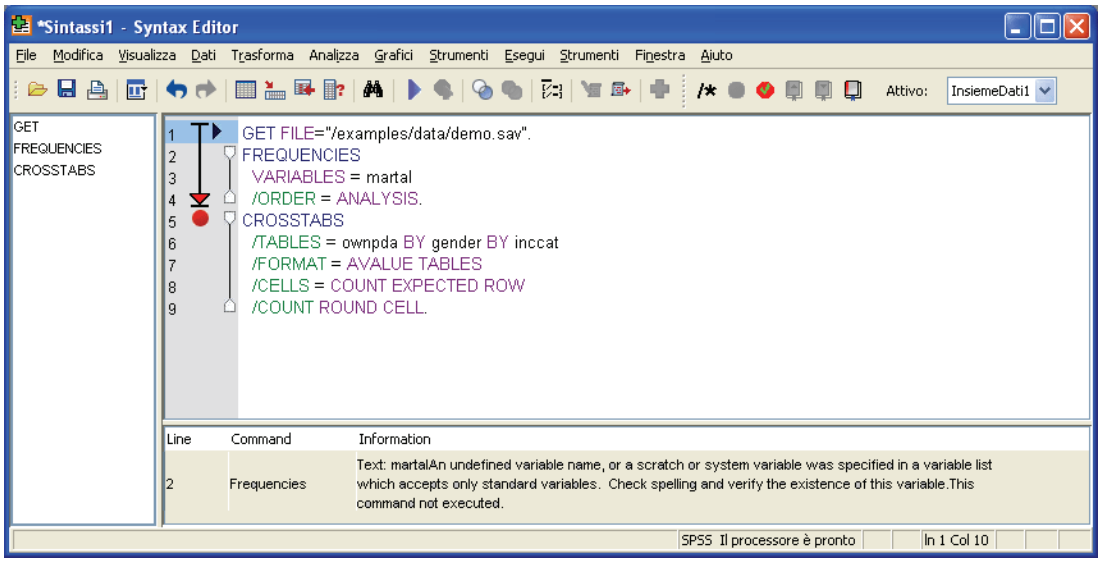
Utilizzo dell'Editor della sintassi

L'Editor della sintassi offre un ambiente ideato appositamente per la creazione, la modifica e l'esecuzione della sintassi dei comandi. L'Editor della sintassi offre le funzionalità seguenti:

- **Autocompletamento.** Durante la digitazione, è possibile selezionare comandi, sottocomandi, parole chiave e valori di parole chiave da un elenco sensibile al contesto. È possibile scegliere che l'elenco venga visualizzato automaticamente oppure su richiesta.
- **Codifica colori.** Gli elementi riconosciuti della sintassi dei comandi (comandi, sottocomandi, parole chiave e valori di parole chiave) sono codificati tramite il colore, permettendo così di vedere immediatamente i termini non riconosciuti. Inoltre, vengono codificati tramite il colore anche un certo numero di errori comuni di sintassi, ad esempio le virgolette non corrispondenti, per una più rapida identificazione.
- **Punti di interruzione.** È possibile interrompere l'esecuzione della sintassi dei comandi in punti specifici, in modo da ispezionare i dati o l'output prima di procedere.
- **Segnalibri.** È possibile impostare segnalibri che permettono di navigare rapidamente all'interno di grandi file di sintassi dei comandi.
- **Esegui.** È possibile eseguire la sintassi di un comando un comando alla volta, passando al comando successivo con un solo clic.

Finestra Editor della sintassi

Figura 13-3
Editor della sintassi


*Sintassi1 - Syntax Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Esegui Strumenti Finestra Aiuto
Attivo: InsiemeDati1
GET
FREQUENCIES
CROSSTABS
1 GET FILE="/examples/data/demo.sav".
2 FREQUENCIES
3 VARIABLES = martal
4 /ORDER = ANALYSIS.
5 CROSSTABS
6 /TABLES = ownpda BY gender BY inccat
7 /FORMAT = AVALUE TABLES
8 /CELLS = COUNT EXPECTED ROW
9 /COUNT ROUND CELL.
Line Command Information
2 Frequencies Text: martalAn undefined variable name, or a scratch or system variable was specified in a variable list which accepts only standard variables. Check spelling and verify the existence of this variable.This command not executed.
SPSS Il processore è pronto
In 1 Col 10

La finestra Editor della sintassi è divisa in quattro sezioni:

- Il riquadro dell'editor è la parte principale della finestra dell'Editor della sintassi e la parte in cui si inserisce e si modifica la sintassi dei comandi.
- La rilegatura è adiacente al riquadro dell'editor e visualizza informazioni quali i numeri di riga e le posizioni dei punti di interruzione.
- Il riquadro di navigazione si trova alla sinistra della rilegatura e del riquadro dell'editor e visualizza un elenco di tutti i comandi nella finestra Editor della sintassi, oltre a consentire una navigazione con un solo clic verso qualsiasi comando.
- Il riquadro degli errori si trova sotto il riquadro dell'editor e visualizza gli errori di runtime.

Contenuti della rilegatura

I numeri di riga, i punti di interruzione, i segnalibri, gli indicatori visivi comandi e un indicatore di avanzamento sono visualizzati nella rilegatura alla sinistra del riquadro dell'editor nella finestra di sintassi.

- I numeri di riga non considerano i file esterni a cui si fa riferimento nei comandi `INSERT` e `INCLUDE`. È possibile mostrare o nascondere i numeri di riga scegliendo Visualizza > Mostra numeri di riga dai menu.
- I punti di interruzione interrompono l'esecuzione in punti specifici e sono rappresentati da un cerchio rosso accanto al comando su cui è impostato il punto di interruzione.
- I segnalibri contrassegnano delle righe specifiche in un file di sintassi dei comandi e sono rappresentati da un quadratino che racchiude il numero (1-9) assegnato al segnalibro. Passando il mouse sull'icona di un segnalibro, vengono visualizzati il numero e il nome del segnalibro, se presenti, assegnati a quel segnalibro.

- Gli indicatori visivi comandi sono icone che offrono un'indicazione visiva dell'inizio e della fine di un comando. È possibile mostrare o nascondere gli indicatori visivi comandi scegliendo Visualizza > Mostra indicatori visivi comandi dai menu.
- L'avanzamento di una determinata esecuzione di sintassi è indicato da una freccia che punta verso il basso nella rilegatura e che si estende dall'esecuzione del primo comando all'esecuzione dell'ultimo comando. Questo indicatore è il più utile quando si esegue la sintassi dei comandi che contiene punti di interruzione e quando si esegue la sintassi dei comandi un'istruzione alla volta. Per ulteriori informazioni, vedere Esecuzione della sintassi dei comandi a pag. 279.

Riquadro di navigazione

Il riquadro di navigazione contiene un elenco di tutti i comandi riconosciuti nella finestra di sintassi, visualizzati nell'ordine in cui si verificano nella finestra. Facendo clic su un comando nel riquadro di navigazione, il cursore viene posizionato all'inizio del comando.

- È possibile utilizzare le frecce su e giù per spostarsi nell'elenco dei comandi o fare clic su un comando per navigare al suo interno. Facendo doppio clic, si seleziona il comando.
- Per impostazione predefinita, i nomi di comando per quei comandi che contengono determinati tipi di errori sintattici, ad esempio virgolette non corrispondenti, sono colorati di rosso con testo in grassetto. Per ulteriori informazioni, vedere Codifica colori a pag. 275.
- La prima parola di ciascuna riga di testo non riconosciuto viene visualizzata in grigio.
- È possibile mostrare o nascondere il riquadro di navigazione scegliendo Visualizza > Mostra riquadro di spostamento dai menu.

Riquadro degli errori

Il riquadro degli errori visualizza gli errori di runtime dall'esecuzione più vecchia.

- Le informazioni su ciascun errore contengono il numero di riga in cui si è verificato l'errore.
- È possibile utilizzare le frecce su e giù per spostarsi all'interno dell'elenco degli errori.
- Facendo clic su una voce dell'elenco, il cursore verrà posizionato sulla riga che ha generato l'errore.
- È possibile mostrare o nascondere il riquadro degli errori scegliendo Visualizza > Mostra riquadro degli errori dai menu.

Terminologia

Comandi. L'unità di base della sintassi è il comando. Ciascun comando inizia con il nome del comando che contiene uno, due o tre parole, ad esempio `DESCRIPTIVES`, `SORT CASES` o `ADD VALUE LABELS`.

Sottocomandi. La maggior parte dei comandi contiene dei sottocomandi. I sottocomandi forniscono specifiche aggiuntive e iniziano con una barra (/) seguita dal nome del sottocomando.

Parole chiave. Le parole chiave sono termini fissi usati solitamente all'interno di un sottocomando per specificare le opzioni disponibili per il sottocomando.

Valori di parole chiave. Le parole chiave possono presentare dei valori, ad esempio un termine fisso che specifica un'opzione o un valore numerico.

Esempio

```
CODEBOOK gender jobcat salary
  /VARINFO VALUELABELS MISSING
  /OPTIONS VARORDER=MEASURE.
```

- Il nome del comando è `CODEBOOK`.
- `VARINFO` e `OPTIONS` sono sottocomandi.
- `VALUELABELS`, `MISSING` e `VARORDER` sono le parole chiave.
- `MEASURE` è un valore di parola chiave associato a `VARORDER`.

Autocompletamento

L'Editor della sintassi fornisce assistenza sottoforma di autocompletamento dei comandi, sottocomandi, parole chiave e valori di parole chiave. Per impostazione predefinita, viene automaticamente visualizzato un elenco sensibile al contesto contenente i termini disponibili. È possibile visualizzare l'elenco su richiesta premendo Ctrl+barra spaziatrice e chiuderlo premendo il tasto Esc.

La voce di menu Completa automaticamente nel menu Strumenti attiva/disattiva la visualizzazione automatica dell'elenco di autocompletamento. È anche possibile abilitare o disabilitare la visualizzazione automatica dell'elenco dalla scheda Editor della sintassi nella finestra di dialogo Opzioni. L'attivazione/disattivazione della voce di menu Completa automaticamente ha la precedenza sull'impostazione nella finestra di dialogo Opzioni, ma non viene mantenuta tra le sessioni.

Nota: l'elenco di autocompletamento verrà chiuso se viene inserito uno spazio. Per i comandi formati da più parole, ad esempio `ADD FILES`, selezionare il comando desiderato prima di inserire qualsiasi spazio.

Codifica colori

L'Editor della sintassi codifica tramite i colori gli elementi riconosciuti della sintassi dei comandi, ad esempio i comandi e i sottocomandi, nonché vari errori sintattici come le virgolette o le parentesi non corrispondenti. Il testo non riconosciuto non viene codificato tramite i colori.

Comandi. Per impostazione predefinita, i comandi riconosciuti sono in blu con testo in grassetto. Se, tuttavia, il comando contiene un errore sintattico riconosciuto, ad esempio una parentesi mancante, per impostazione predefinita il nome del comando viene colorato in rosso con testo in grassetto.

Nota: le abbreviazioni dei nomi di comandi, ad esempio `FREQ` per `FREQUENCIES`, non sono colorate, ma sono comunque valide.

Sottocomandi. Per impostazione predefinita i sottocomandi riconosciuti sono colorati in verde. Se, tuttavia, nel sottocomando manca un segno di uguale obbligatorio oppure se il sottocomando è seguito da un segno di uguale non valido, il nome del sottocomando viene colorato in rosso per impostazione predefinita.

Parole chiave. Per impostazione predefinita le parole chiave riconosciute sono colorate in marrone. Se, tuttavia, nella parola chiave manca un segno di uguale obbligatorio oppure se la parola chiave è seguita da un segno di uguale non valido, la parola chiave viene colorata in rosso per impostazione predefinita.

Valori di parole chiave. Per impostazione predefinita i valori di parola chiave riconosciuti sono colorati in rosa. I valori di parole chiave definiti dall'utente, ad esempio interi, numeri reali e stringhe tra virgolette, non sono codificati tramite i colori.

Commenti. Il testo all'interno di un commento viene colorato in grigio.

Errori sintattici. Per impostazione predefinita, il testo associato ai seguenti errori sintattici viene colorato in rosso.

- **Parentesi tonde, parentesi quadre e virgolette non accoppiate.** Le parentesi tonde e le parentesi quadre non accoppiate all'interno dei commenti non vengono rilevate, così come le stringhe tra virgolette. Le virgolette singole o doppie non accoppiate all'interno delle stringhe tra virgolette sono sintatticamente valide.

  Alcuni comandi contengono blocchi di testo che non sono sintassi di comando, ad esempio `BEGIN DATA-END DATA`, `BEGIN GPL-END GPL` e `BEGIN PROGRAM-END PROGRAM`. I valori non corrispondenti non vengono rilevati all'interno di tali blocchi.
- **Righe lunghe.** Le righe lunghe sono righe contenenti più di 251 caratteri.
- **Istruzioni End.** Molti comandi richiedono o un'istruzione `END` prima del terminatore di comando (ad esempio, `BEGIN DATA-END DATA`) oppure richiedono un comando `END` corrispondente in un punto successivo del flusso di comando (ad esempio, `LOOP-END LOOP`). In entrambi i casi, il comando verrà colorato in rosso per impostazione predefinita, finché non viene aggiunta l'istruzione `END` richiesta.

Dalla scheda Editor della sintassi nella finestra di dialogo Opzioni, è possibile modificare i colori e gli stili di testo predefiniti, nonché attivare o disattivare la codifica colori. È possibile attivare o disattivare la codifica colori di comandi, sottocomandi, parole chiave e valori di parole chiave anche da Strumenti > Codifica colori. È possibile attivare o disattivare la codifica colori degli errori sintattici da Strumenti > Convalida. Le scelte effettuate nel menu Strumenti hanno la precedenza sulle impostazioni della finestra di dialogo Opzioni, ma non vengono mantenute tra le varie sessioni.

Nota: la codifica colori della sintassi dei comandi all'interno delle macro non è supportata.

Punti di interruzione

I punti di interruzione consentono di interrompere l'esecuzione della sintassi dei comandi in punti specifici all'interno della finestra di sintassi e di riprendere l'esecuzione quando si è pronti.

- I punti di interruzione sono impostati al livello di un comando e interrompono l'esecuzione prima di eseguire il comando.

- I punti di interruzione non possono essere inseriti all'interno dei blocchi `LOOP-END LOOP`, `DO IF-END IF` , `DO REPEAT-END REPEAT`, `INPUT PROGRAM-END INPUT PROGRAM` e `MATRIX-END MATRIX`. Tuttavia, possono essere impostati all'inizio di tali blocchi e interromperanno l'esecuzione prima di eseguire il blocco.
- I punti di interruzione non possono essere impostati sulle righe che contengono sintassi di comando non-SPSS Statistics, come avviene nei blocchi `BEGIN PROGRAM-END PROGRAM`, `BEGIN DATA-END DATA` e `BEGIN GPL-END GPL`.
- I punti di interruzione non vengono salvati con il file di sintassi dei comandi e non sono inclusi nel testo copiato.
- Per impostazione predefinita, i punti di interruzione vengono considerati durante l'esecuzione. È possibile decidere se considerare o meno i punti di interruzione da Strumenti > Considera punti di interruzione.

Per inserire un punto di interruzione

▶ Fare clic in un punto qualsiasi della rilegatura alla sinistra del testo del comando.

o

▶ Posizionare il cursore all'interno del comando desiderato.

▶ Dai menu, scegliere:
Strumenti
Attiva/disattiva punto di interruzione

Il punto di interruzione è rappresentato da un cerchio rosso nella rilegatura alla sinistra del testo del comando e sulla stessa riga del nome del comando.

Cancellazione dei punti di interruzione

Per cancellare un solo punto di interruzione:

▶ Fare clic sull'icona che rappresenta il punto di interruzione nella rilegatura alla sinistra del testo del comando.

o

▶ Posizionare il cursore all'interno del comando desiderato.

▶ Dai menu, scegliere:
Strumenti
Attiva/disattiva punto di interruzione

Per cancellare tutti i punti di interruzione:

▶ Dai menu, scegliere:
Strumenti
Cancella tutti i punti di interruzione

Vedere Esecuzione della sintassi dei comandi a pag. 279 per informazioni sul comportamento in fase di runtime in presenza di punti di interruzione.

Segnalibri

I segnalibri consentono di spostarsi rapidamente nelle posizioni specificate in un file di sintassi dei comandi. È possibile avere fino a 9 segnalibri in un determinato file. I segnalibri vengono salvati con il file, ma non inclusi quando si copia il testo.

Per inserire un segnalibro

- Posizionare il cursore in corrispondenza della riga in cui si desidera inserire il segnalibro.
- Dai menu, scegliere:

  Strumenti
  Attiva/Disattiva segnalibro

Il nuovo segnalibro viene assegnato al numero successivo disponibile, da 1 a 9. È rappresentato come un quadrato contenente il numero assegnato ed è visualizzato nella rilegatura alla sinistra del testo del comando.

Cancellazione dei segnalibri

Per cancellare un solo segnalibro:

- Posizionare il cursore sulla riga contenente il segnalibro.
- Dai menu, scegliere:

  Strumenti
  Attiva/Disattiva segnalibro

Per cancellare tutti i segnalibri:

- Dai menu, scegliere:

  Strumenti
  Cancella tutti i segnalibri

Ridenominazione di un segnalibro

È possibile associare un nome a un segnalibro. Questo in aggiunta al numero (1-9) assegnato al segnalibro al momento della creazione.

- Dai menu, scegliere:

  Strumenti
  Rinomina segnalibro
- Inserire un nome per il segnalibro e fare clic su OK.

Il nome specificato sostituisce qualsiasi nome esistente per il segnalibro.

Spostamento tra i segnalibri

Per spostarsi al segnalibro precedente o successivo:

▶ Dai menu, scegliere:

Strumenti
  Segnalibro successivo

o

Strumenti
  Segnalibro precedente

Per spostarsi su un segnalibro specifico:

▶ Dai menu, scegliere:

Strumenti
  Vai al segnalibro

▶ Selezionare il segnalibro desiderato.

Impostazione del testo come commento

È possibile impostare come commento interi comandi, nonché il testo che non è riconosciuto come sintassi dei comandi.

▶ Selezionare il testo desiderato. Quando si selezionano più comandi, un comando verrà impostato come commento se è selezionata una sua parte qualsiasi.

▶ Dai menu, scegliere:

Strumenti
  Selezione commento

È possibile impostare come commento un singolo comando posizionando il cursore in qualsiasi punto all'interno del comando e facendo clic sul pulsante Selezione commento nella barra degli strumenti.

Esecuzione della sintassi dei comandi

▶ Evidenziare i comandi che si desidera eseguire nella finestra della sintassi.

▶ Fare clic sul pulsante Esegui (il triangolo rivolto verso destra) nella barra degli strumenti dell'Editor della sintassi. Vengono eseguiti i comandi selezionati, oppure se non è presente alcuna selezione, viene eseguito il comando in cui è posizionato il cursore.

o

▶ Scegliere una delle voci dal menu Esegui.

- **Tutto.** Vengono eseguiti tutti i comandi presenti nella finestra di sintassi, considerando qualsiasi punto di interruzione.
- **Selezione.** Vengono eseguiti solo i comandi correntemente selezionati, considerando qualsiasi punto di interruzione. Sono compresi i comandi parzialmente evidenziati. Se non è stata effettuata alcuna selezione, viene eseguito il comando in cui è posizionato il cursore.

- **Fino alla fine.** Vengono eseguiti tutti i comandi iniziando dal primo comando nella selezione corrente fino all'ultimo comando nella finestra della sintassi, considerando qualsiasi punto di interruzione. Se non è presente alcuna selezione, l'esecuzione parte dal comando in cui è posizionato il cursore.
- **Esegui.** Viene eseguita la sintassi di comando un comando alla volta iniziando dal primo comando presente nella finestra di sintassi (Esegui dall'inizio) oppure dal comando in cui è posizionato il cursore (Esegui dal corrente). Se è presente del testo selezionato, l'esecuzione inizia dal primo comando nella selezione. Dopo che è stato eseguito un determinato comando, il cursore avanza al comando successivo e si procede nell'esecuzione scegliendo Continua.

  Quando si utilizza Esegui, i blocchi `LOOP-END LOOP`, `DO IF-END IF`, `DO REPEAT-END REPEAT`, `INPUT PROGRAM-END INPUT PROGRAM` e `MATRIX-END MATRIX` vengono trattati come comandi singoli. Non è possibile eseguire l'istruzione di uno di questi blocchi.
- **Continua.** Continua un'esecuzione interrotta da un punto di interruzione o da Esegui.

Indicatore di avanzamento

L'avanzamento dell'esecuzione di una sintassi data viene indicato con una freccia che punta verso il basso nella rilegatura, estendendosi sull'ultimo insieme di comandi eseguiti. Ad esempio, si decide di eseguire tutti i comandi in una finestra di sintassi che contiene punti di interruzione. Al primo punto di interruzione, la freccia estenderà l'area dal primo comando nella finestra al comando precedente a quello che contiene il punto di interruzione. Al secondo punto di interruzione, la freccia si allungherà dal comando che contiene il primo punto di interruzione fino al comando precedente a quello che contiene il secondo punto di interruzione.

Comportamento in fase di runtime con i punti di interruzione

- Quando si esegue la sintassi di comando che contiene punti di interruzione, l'esecuzione viene interrotta a ciascun punto di interruzione. Nello specifico, il blocco della sintassi di comando da un punto di interruzione dato (o inizio dell'esecuzione) al punto di interruzione successivo (o fine dell'esecuzione) viene inviato per l'esecuzione esattamente come se si fosse selezionata quella sintassi e si fosse scelto Esegui > Selezione.
- È possibile lavorare con più finestre di sintassi, ciascuna con il proprio insieme di punti di interruzione, ma esiste una sola coda per eseguire la sintassi dei comandi. Dopo aver inviato un blocco di sintassi dei comandi, ad esempio un blocco di sintassi dei comandi fino al primo punto di interruzione, nessun altro blocco di sintassi dei comandi verrà eseguito fino al completamente del blocco precedente, indipendentemente dal fatto che i blocchi siano nella stessa finestra di sintassi o in finestre separate.
- Quando l'esecuzione è interrotta in corrispondenza di un punto di interruzione, è possibile eseguire la sintassi dei comandi in altre finestre di sintassi e ispezionare le finestre dell'Editor dei dati o del Viewer. Tuttavia, la modifica dei contenuti della finestra di sintassi che contiene il punto di interruzione o la modifica della posizione del cursore in quella finestra, causerà l'annullamento dell'esecuzione.

File di sintassi Unicode

In modalità Unicode, il formato predefinito per il salvataggio dei file di sintassi dei comandi creati o modificati durante la sessione è anche Unicode (UTF-8). I file di sintassi dei comandi in formato Unicode non possono essere letti dalle versioni di SPSS Statistics precedenti alla 16.0. Per ulteriori informazioni sulla modalità Unicode, vedere Opzioni generali a pag. 522.

Per salvare un file di sintassi in un formato compatibile con le versioni precedenti:

- Dai menu della finestra di sintassi, scegliere:
  File
  Salva con nome
- Nell'elenco a discesa Codifica della finestra di dialogo Salva con nome, scegliere Codifica locale. La codifica locale viene determinata dalle impostazioni internazionali correnti.

Comandi EXECUTE multipli

La sintassi copiata dalle finestre di dialogo o copiata dal registro o dal giornale potrebbe contenere i comandi EXECUTE. Se si eseguono più comandi da una finestra di sintassi, di norma non sono necessari i comandi EXECUTE che potrebbero ridurre le prestazioni, specie con file di grandi dimensioni, perché tale comando legge tutto il file di dati. Per ulteriori informazioni, vedere il comando EXECUTE in *Command Syntax Reference* (disponibile tramite il menu ? in qualsiasi finestra di SPSS Statistics).

Ritardi

Un'eccezione interessante è rappresentata dai comandi di trasformazione che contengono i ritardi. In una serie di comandi di trasformazione senza l'impiego dei comandi EXECUTE o di altri comandi per la lettura dei dati, i ritardi vengono calcolati dopo tutte le altre trasformazioni, indipendentemente dall'ordine dei comandi. Ad esempio,

```
COMPUTE lagvar=LAG(var1).
COMPUTE var1=var1*2.
```

e

```
COMPUTE lagvar=LAG(var1).
EXECUTE.
COMPUTE var1=var1*2.
```

Generano risultati molto diversi per il valore di *lagvar*, poiché il primo comando utilizza il valore trasformato di *var1* mentre il secondo usa il valore originale.

Capitolo
14

Informazioni sui dati

Le informazioni sui dati restituiscono informazioni del dizionario, ad esempio nomi di variabili, etichette di variabili e di valori o valori mancanti, e statistiche riassuntive per alcune o tutte le variabili e gli insiemi a risposta multipla presenti nell'insieme di dati attivo. Per le variabili nominali e ordinali e gli insiemi a risposta multipla, le statistiche riassuntive includono conteggi e percentuali. Per le variabili di scala, le statistiche riassuntive includono media, deviazione standard e quartili.

Nota: Informazioni sui dati ignorano lo stato del file distinto. Sono inclusi i gruppi di file distinti creati per i valori mancanti ad assegnazione multipla (disponibili nel modulo aggiuntivo Missing Values). Per ulteriori informazioni sull'elaborazione dei file distinti, vedere Dividi.

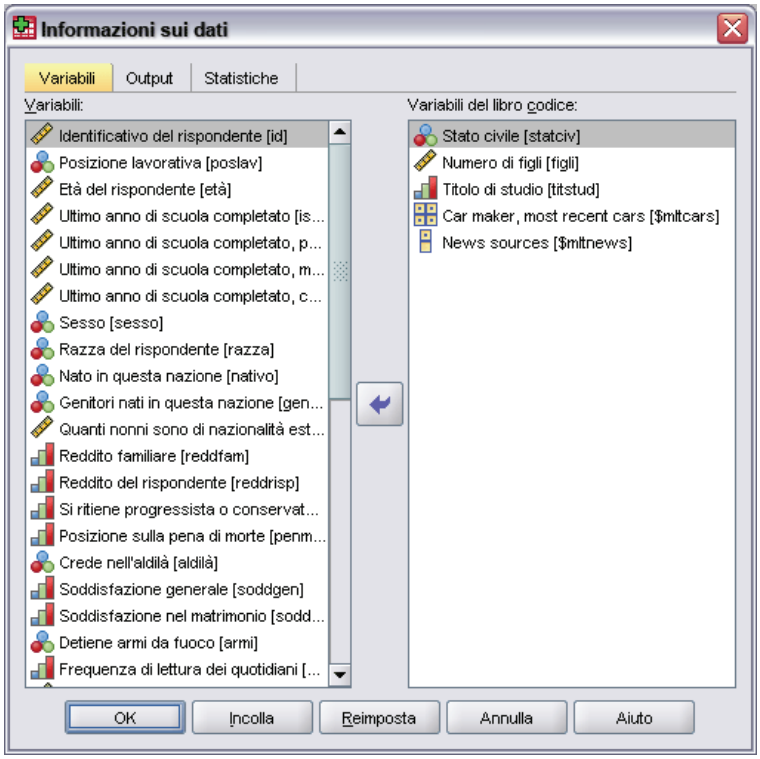
Per ottenere le informazioni sui dati

- Dai menu, scegliere:

  Analizza
  Report
  Informazioni sui dati

- Fare clic sulla scheda Variabili.

Figura 14-1
Finestra di dialogo Informazioni sui dati, scheda Variabili



- Selezionare una o più variabili e/o uno o più insiemi a risposta multipla.

Se lo si desidera, è possibile:

- Controllare le informazioni sulle variabili visualizzate.
- Controllare le statistiche visualizzate (o escludere tutte le statistiche riassuntive).
- Controllare l'ordine in cui vengono visualizzati insiemi a risposta multipla e variabili.
- Modificare il livello di misurazione per le variabili nell'elenco di origine in modo tale da modificare le statistiche riassuntive visualizzate. Per ulteriori informazioni, vedere Scheda Informazioni sui dati - Statistiche a pag. 287.

Modifica del livello di misurazione

È possibile modificare temporaneamente il livello di misurazione per le variabili. Non è possibile modificare il livello di misurazione per gli insiemi a risposta multipla, che vengono sempre trattati come nominali)

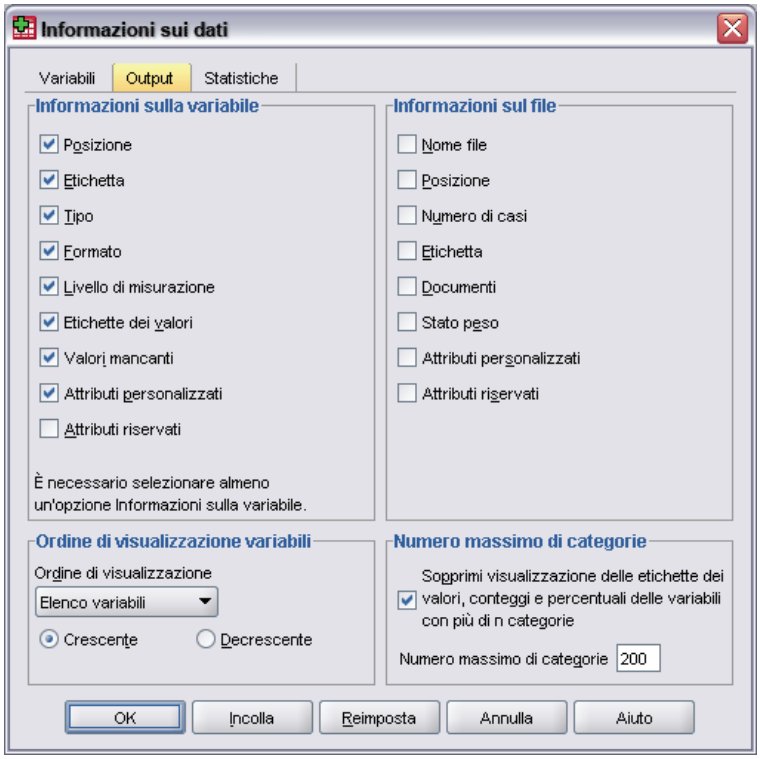
- Fare clic con il pulsante destro del mouse su una variabile nell'elenco sorgente.
- Scegliere un livello di misurazione dal menu di scelta rapida popup.

In questo modo il livello di misurazione viene temporaneamente modificato. Da un punto di vista pratico, è utile solo per le variabili numeriche. Il livello di misurazione per le variabili stringa è limitato a nominale o ordinale, entrambi trattati allo stesso modo dalla procedura Informazioni sui dati.

Scheda Output della finestra Informazioni sui dati

La scheda Output controlla le informazioni sulle variabili incluse per ciascuna variabile e ciascun insieme a risposta multipla, l'ordine in cui variabili e insiemi a risposta multipla vengono visualizzati e il contenuto della tabella delle informazioni facoltative sui file.

Figura 14-2
Finestra di dialogo Informazioni sui dati, scheda Output



Informazioni sulla variabile

Controlla le informazioni del dizionario visualizzate per ciascuna variabile.

Posizione. Intero che rappresenta la posizione della variabile nell'ordine del file. Non è disponibile per gli insiemi a risposta multipla.

Etichetta. Etichetta descrittiva associata alla variabile o all'insieme a risposta multipla.

Tipo. Tipo di dati fondamentale. Può essere *Numerico*, *Stringa* o *Insieme a risposta multipla*.

Formato. Formato di visualizzazione della variabile, ad esempio *A4*, *F8.2* o *DATE11*. Non è disponibile per gli insiemi a risposta multipla.

Livello di misurazione. I valori possibili sono *Nominale*, *Ordinale*, *Scala* e *Sconosciuto*. Il valore visualizzato è il livello di misurazione memorizzato nel dizionario e non è influenzato da alcuna variazione temporanea del livello di misurazione dovuta alla modifica del livello nell'elenco delle variabili sorgente della scheda Variabili. Non è disponibile per gli insiemi a risposta multipla.

Nota: il livello di misurazione per le variabili numeriche può essere "sconosciuto" prima del primo ciclo di dati poiché tale livello non è stato ancora impostato esplicitamente, ad esempio quando i dati vengono letti da una sorgente esterna o le variabili sono appena state create.Per ulteriori informazioni, vedere Livello di misurazione delle variabili in Capitolo 5 a pag. 80.

Etichette dei valori. Etichette descrittive associate a valori di dati specifici. Per ulteriori informazioni, vedere Etichette dei valori in Capitolo 5 a pag. 84.

- Se nella scheda Statistiche è selezionata l'opzione Conteggio o Percentuale, le etichette dei valori definiti sono incluse nell'output anche se l'opzione Etichette dei valori non è stata selezionata.
- Per gli insiemi a dicotomie multiple, le "etichette dei valori" sono le etichette delle variabili per le variabili elementari presenti nell'insieme o le etichette dei valori conteggiati, in base alla definizione dell'insieme. Per ulteriori informazioni, vedere Insiemi a risposta multipla in Capitolo 7 a pag. 114.

Valori mancanti. Valori mancanti definiti dall'utente. Per ulteriori informazioni, vedere Valori mancanti in Capitolo 5 a pag. 85. Se nella scheda Statistiche è selezionata l'opzione Conteggio o Percentuale, le etichette dei valori definiti sono incluse nell'output anche se l'opzione Valori mancanti non è stata selezionata. Non è disponibile per gli insiemi a risposta multipla.

Attributi personalizzati. Attributi delle variabili definite dall'utente. L'output include sia i nomi sia i valori di tutti gli attributi personalizzati associati a ciascuna variabile. Per ulteriori informazioni, vedere Creazione di attributi di variabile personalizzati in Capitolo 5 a pag. 88. Non è disponibile per gli insiemi a risposta multipla.

Attributi riservati. Attributi riservati delle variabili di sistema. È possibile visualizzare gli attributi di sistema, ma non modificarli. I nomi degli attributi di sistema iniziano con il simbolo del dollaro ($) . Gli attributi non di visualizzazione, che hanno nomi che iniziano con "@" o "$@", non sono inclusi. L'output include sia i nomi sia i valori di tutti gli attributi di sistema associati a ciascuna variabile. Non è disponibile per gli insiemi a risposta multipla.

Informazioni sul file

La tabella delle informazioni opzionali sul file può includere i seguenti attributi:

Nome file. Nome del file di dati di SPSS Statistics. Se l'insieme di dati non è mai stato salvato in formato SPSS Statistics, non esiste alcun nome di file di dati. Se non è visualizzato un nome di file nella barra del titolo della finestra Editor dei dati, l'insieme di dati attivo non ha un nome di file.

Posizione. Percorso della directory (cartella) del file di dati di SPSS Statistics. Se l'insieme di dati non è mai stato salvato in formato SPSS Statistics, non esiste alcun percorso.

Numero di casi. Numero di casi presenti nell'insieme di dati attivo. Si tratta del numero totale di casi, compresi gli eventuali casi esclusi dalle statistiche riassuntive a causa delle condizioni di filtro.

Etichetta. Etichetta del file (se presente) definita dal comando `FILE LABEL`.

Documenti. Testo del documento del file di dati. Per ulteriori informazioni, vedere Commenti sul file di dati in Capitolo 47 a pag. 517.

Stato della ponderazione. Se il peso è attivo, viene visualizzato il nome della variabile di peso. Per ulteriori informazioni, vedere Pesa casi in Capitolo 9 a pag. 197.

Attributi personalizzati. Attributi del file di dati definiti dall'utente. Gli attributi del file di dati vengono definiti con il comando `DATAFILE ATTRIBUTE`.

Attributi riservati. Attributi riservati del file di dati di sistema. È possibile visualizzare gli attributi di sistema, ma non modificarli. I nomi degli attributi di sistema iniziano con il simbolo del dollaro ($) . Gli attributi non di visualizzazione, che hanno nomi che iniziano con "@" o "$@", non sono inclusi. L'output include sia i nomi che i valori di tutti gli attributi del file di dati di sistema.

Ordine di visualizzazione variabili

Sono disponibili le seguenti opzioni per il controllo dell'ordine in cui vengono visualizzati gli insiemi a risposta multipla e le variabili.

Alfabetico. Ordine alfabetico per nome di variabile.

File. Ordine in cui le variabili appaiono nell'insieme di dati (l'ordine in cui sono visualizzate nell'Editor dei dati). In ordine crescente, con gli insiemi a risposta multipla visualizzati per ultimi, dopo tutte le variabili selezionate.

Livello di misurazione. Ordina per livello di misurazione. Vengono creati quattro gruppi di ordinamento: nominale, ordinale, scala e sconosciuto. Gli insiemi a risposta multipla vengono trattati come nominali.

Nota: il livello di misurazione per le variabili numeriche può essere "sconosciuto" prima del primo ciclo di dati poiché tale livello non è stato ancora impostato esplicitamente, ad esempio quando i dati vengono letti da una sorgente esterna o le variabili sono appena state create.

Elenco variabili. Ordine in cui le variabili e gli insiemi a risposta multipla appaiono nell'elenco delle variabili selezionate della scheda Variabili.

Nome attributo personalizzato. L'elenco delle opzioni di ordinamento include anche i nomi degli attributi personalizzati delle variabili definite dall'utente. In ordine crescente, con le variabili senza attributi per prime, seguite dalle variabili che hanno un attributo ma senza valori definiti per lo stesso, seguite dalle variabili con valori definiti per l'attributo e valori in ordine alfabetico.

Numero massimo di categorie

Se l'output include etichette di valori, conteggi o percentuali per ogni valore univoco, è possibile eliminare questa informazione dalla tabella se il numero di valori supera il valore specificato. Per impostazione predefinita, questa informazione non viene inserita se il numero di valori univoci per la variabile supera 200.

Scheda Informazioni sui dati - Statistiche

La scheda Statistiche consente di controllare le statistiche riassuntive incluse nell'output o di eliminare completamente la visualizzazione delle statistiche riassuntive.

Figura 14-3
Finestra di dialogo Informazioni sui dati, scheda Statistiche

Informazioni sui dati
Variabili
Output
Statistiche
Conteggi e percentuali
Conteggi
Percentuale
I conteggi e le percentuali si applicano a tutte le variabili nominali e ordinali (incluse tutte le variabili stringa), agli insiemi a risposta multipla e ai valori con etichetta delle variabili di scala.
Tendenza centrale e dispersione
Media
Deviazione standard
Quartili
Le misure della tendenza centrale e della dispersione si applicano solo alle variabili di scala.
OK
Incolla
Reimposta
Annulla
Aiuto

Conteggi e percentuali

Per le variabili nominali e ordinali, gli insiemi a risposta multipla e i valori o le variabili di scala con etichetta, le statistiche riassuntive sono:

Conteggio. Il conteggio o numero di casi con ogni valore (o intervallo di valori) di una variabile.

Percentuale. La percentuale di casi con un valore particolare.

Tendenza centrale e dispersione

Per le variabili di scala, le statistiche disponibili sono:

Media. Una misura di tendenza centrale. La somma dei valori di tutte le osservazioni divisa per il numero di osservazioni. Viene anche detta media aritmetica.

Deviazione standard. La radice quadrata della varianza. La deviazione standard è una misura della dispersione intorno alla media espressa nella stessa unità di misura delle osservazioni. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95%

dei casi rientra in due deviazioni standard. Se, ad esempio, in una popolazione con distribuzione normale l'età media fosse 45 e la deviazione standard 10, il 95% dei casi cadrebbe fra 25 e 65 anni.

Quartili. Mostra i valori corrispondenti al 25°, 50° e 75° percentile.

Nota: è possibile modificare temporaneamente il livello di misurazione associato a una variabile (e quindi modificare le statistiche riassuntive visualizzate per quella variabile) nell'elenco variabili sorgente della scheda Variabili.

Capitolo
15

Frequenze

La procedura Frequenze consente di ottenere statistiche e rappresentazioni grafiche che risultano utili per la descrizione di molti tipi di variabili. La procedura Frequenza offre un'ottima opportunità per iniziare ad osservare i dati.

Per ottenere un rapporto e un grafico a barre delle frequenze è possibile disporre i singoli valori in ordine crescente o decrescente oppure ordinare le categorie in base alle rispettive frequenze. Il rapporto sulle frequenze può essere eliminato se una variabile ha molti valori distinti. È possibile etichettare i grafici con frequenze (default) o percentuali.

Esempio. Qual è la distribuzione dei clienti di un'azienda per tipo di industria? Dall'output, si nota che il 37,5% dei clienti fa parte di enti governativi, il 24,9% fa parte di società, il 28,1% di istituzioni accademiche e il 9,4% del settore sanitario. Per i dati quantitativi e continui, ad esempio il fatturato, si può notare che la vendita media del prodotto è pari a €. 3.576 con una deviazione standard di €. 1.078.

Statistiche e grafici. Conteggi di frequenza, percentuali, percentuali cumulate, media, mediana, moda, somma, deviazione standard, varianza, intervallo, valori minimo e massimo, errore standard della media, asimmetria e curtosi (entrambe con errori standard), quartili, percentili definiti dall'utente, grafici a barre, grafici a torta e istogrammi.

Dati. Utilizzare codici numerici o stringhe per codificare le variabili categoriali (misure di livello nominale o ordinale).

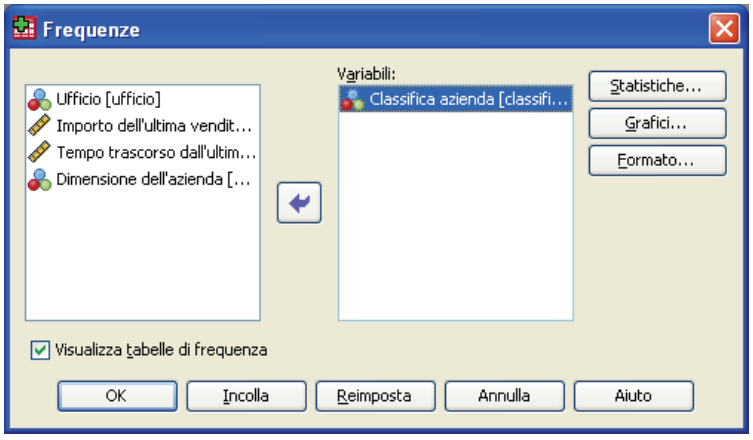
Assunzioni. I riepiloghi e le percentuali forniscono un'utile descrizione dei dati provenienti da qualsiasi distribuzione, in particolare per le variabili con categorie ordinate o non ordinate. La maggior parte delle statistiche riassuntive, ad esempio la media e la deviazione standard, si basano sulla normale teoria e sono idonee per variabili quantitative con distribuzioni simmetriche. Le statistiche robuste, ad esempio la media, i quartili e i percentili, sono idonee per variabili quantitative rispondenti o meno all'ipotesi di normalità.

Per ottenere le tabelle di frequenza

▶ Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Frequenze...

Figura 15-1
Finestra di dialogo principale Frequenze



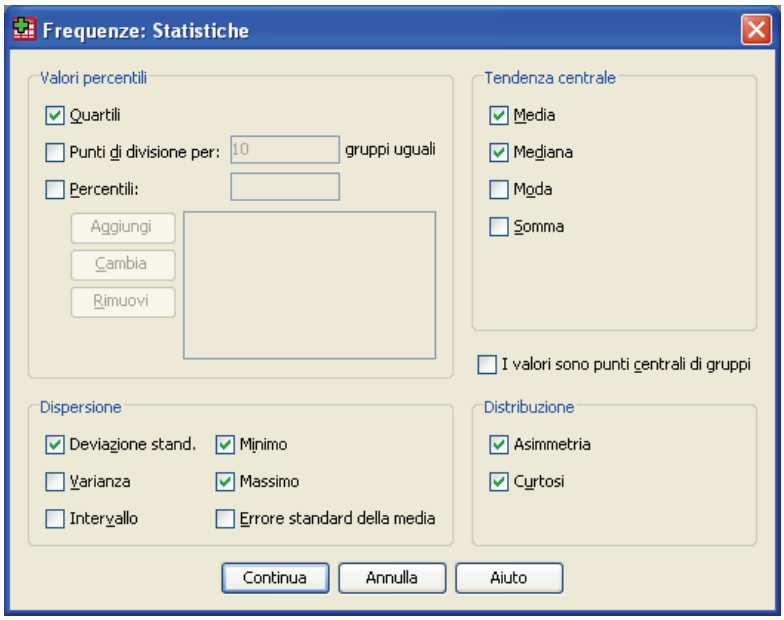
- Selezionare una o più variabili categoriali o quantitative.

Se lo si desidera, è possibile:

- Fare clic su Statistiche per ottenere statistiche descrittive per le variabili quantitative.
- Fare clic su Grafici per ottenere grafici a barre, grafici a torta e istogrammi.
- Fare clic su Formato per stabilire l'ordine in cui visualizzare i risultati.

Frequenze: Statistiche

Figura 15-2
Finestra di dialogo Frequenze: Statistiche



Valori percentili. Valori di una variabile quantitativa che suddividono i dati ordinati in due gruppi in modo da visualizzare una percentuale sopra e una sotto. I quartili (il 25°, 50° e 75° percentile) suddividono le osservazioni in quattro gruppi di dimensioni uguali. Se si desidera ottenere un numero di gruppi uguali diverso da quattro, selezionare Punti di divisione per gruppi uguali. È inoltre possibile specificare i singoli percentili, ad esempio il 95° percentile, ovvero il valore al di sotto del quale ricade il 95% delle osservazioni.

Tendenza centrale. Le statistiche che descrivono la posizione della distribuzione includono media, mediana, moda e somma di tutti i valori.

- **Media.** Una misura di tendenza centrale. La somma dei valori di tutte le osservazioni divisa per il numero di osservazioni. Viene anche detta media aritmetica.
- **Mediana.** È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori eccezionalmente bassi o alti.
- **Moda.** Il valore che ricorre più frequentemente. Se più valori condividono la maggiore ricorrenza, ognuno di essi è una moda. La procedura Frequenze riporta solo la più piccola delle mode.
- **Somma.** La somma o il totale di tutti i valori non mancanti di tutti i casi.

Dispersione. Le statistiche che misurano l'entità della variazione o della variabilità dei dati includono deviazione standard, varianza, intervallo, valore minimo e massimo ed errore standard della media.

- **Deviazione stand..** La radice quadrata della varianza. La deviazione standard è una misura della dispersione intorno alla media espressa nella stessa unità di misura delle osservazioni. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, in una popolazione con distribuzione normale l'età media fosse 45 e la deviazione standard 10, il 95% dei casi cadrebbe fra 25 e 65 anni.
- **Varianza.** Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile.
- **Intervallo.** La differenza tra il valore massimo ed il valore minimo di una variabile numerica.
- **Minimo.** Il valore più basso assunto da una variabile numerica.
- **Massimo.** Il valore più alto di una variabile numerica.
- **E. S. media.** Una misura di quanto può variare il valore della media da campione a campione per campioni estratti dalla stessa distribuzione. Può essere utilizzata per confrontare genericamente la media osservata rispetto a un valore ipotizzato (ovvero, è possibile concludere che i due valori sono diversi se il rapporto della differenza rispetto all'errore standard è inferiore a -2 o maggiore di +2).

Distribuzione. L'asimmetria e la curtosi sono statistiche che descrivono la forma e la simmetria della distribuzione. Queste statistiche vengono visualizzate con i relativi errori standard.

- **Asimmetria.** Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con una notevole asimmetria positiva ha una lunga coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica lo scostamento dalla normale simmetria.
- **Curtosi.** Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per la distribuzione normale, il valore della statistica di curtosi è zero. La curtosi positiva indica che le osservazioni sono più raggruppate e hanno una coda più lunga rispetto a quelle della distribuzione normale. La curtosi negativa indica che le osservazioni sono meno raggruppate e hanno code più corte.

I valori sono punti centrali di gruppi. Se i valori dei dati sono punti centrali di gruppi (ad esempio, l'età delle persone sulla trentina è codificata come 35), selezionare questa opzione per valutare la media e i percentili per i dati originali non raggruppati.

Frequenze: Grafici

Figura 15-3
Finestra di dialogo Frequenze: Grafici.

Tipo di grafico. I grafici a torta mostrano il contributo delle parti all'intero grafico. Ogni sezione di un grafico a torta corrisponde a un gruppo definito da una singola variabile di raggruppamento. Nei grafici a barre il conteggio relativo a ciascun valore o categoria viene rappresentato come una barra distinta, in modo da poter confrontare visivamente le categorie. Anche gli istogrammi contengono barre, che però sono tracciate lungo una scala per intervalli uguali. L'altezza di ogni barra rappresenta il conteggio dei valori di una variabile quantitativa che rientra nell'intervallo. Nell'istogramma vengono indicati la forma, il centro e la variabilità della distribuzione. Una curva normale sovrapposta all'istogramma consente di valutare se i dati sono distribuiti normalmente.

Valori nel grafico. Per i grafici a barre, l'asse di scala può essere etichettato in base ai conteggi o alle percentuali di frequenza.

Frequenze: Formato

Figura 15-4
Finestra di dialogo Frequenze: Formato

Ordina per. La tabella di frequenza può essere disposta in base ai valori effettivi dei dati oppure in base al conteggio (frequenza di ricorrenza) di tali valori, in ordine crescente o decrescente. Se, tuttavia, si desidera ottenere un istogramma o i percentili, si presume che la variabile sia quantitativa e i suoi valori vengano visualizzati in ordine crescente.

Variabili multiple. Se si producono tabelle di statistiche per variabili multiple, è possibile visualizzare tutte le variabili in un'unica tabella (Confronta variabili) o visualizzare una tabella distinta di statistiche per ciascuna variabile (Output per variabili).

Sopprimi le tabelle con più di n categorie. Questa opzione consente di disattivare la visualizzazione delle tabelle che includono un numero di valori maggiore di quello specificato.

Capitolo
16

Descrittive

La procedura Descrittive consente di visualizzare statistiche riassuntive univariate per diverse variabili incluse nella stessa tabella e di calcolare i valori standardizzati (punteggi *z*). È possibile ordinare le variabili in base alle dimensioni delle rispettive medie (in ordine crescente o decrescente), in ordine alfabetico oppure nell'ordine in cui sono state selezionate (impostazione predefinita).

I punteggi *z* salvati vengono aggiunti ai dati nell'Editor dei dati e sono disponibili per la creazione di grafici, elenchi di dati e analisi. Quando le variabili vengono registrate in unità diverse (ad esempio, prodotto interno lordo pro capite e percentuale di alfabetizzazione), una trasformazione dei punteggi *z* consente di posizionare le variabili su una scala comune per facilitarne il confronto visivo.

Esempio. Se ciascun caso incluso nei dati contiene i totali delle vendite giornaliere relativi a ciascun agente di vendita (ad esempio, una voce per Roberto, una per Carlo e una per Bruno), registrati ogni giorno per diversi mesi, la procedura Descrittive consente di calcolare la media delle vendite giornaliere per ogni agente e di ordinare i risultati dalla media di vendita maggiore alla minore.

Statistiche. Dimensioni del campione, media, valore minimo e massimo, deviazione standard, varianza, intervallo, somma, errore standard della media, curtosi e asimmetria degli errori standard.

Dati. Utilizzare variabili numeriche dopo averle valutate graficamente per registrare errori, valori anomali e anomalie distributive. La procedura Descrittive risulta molto utile quando si utilizzano file di grandi dimensioni (migliaia di casi).

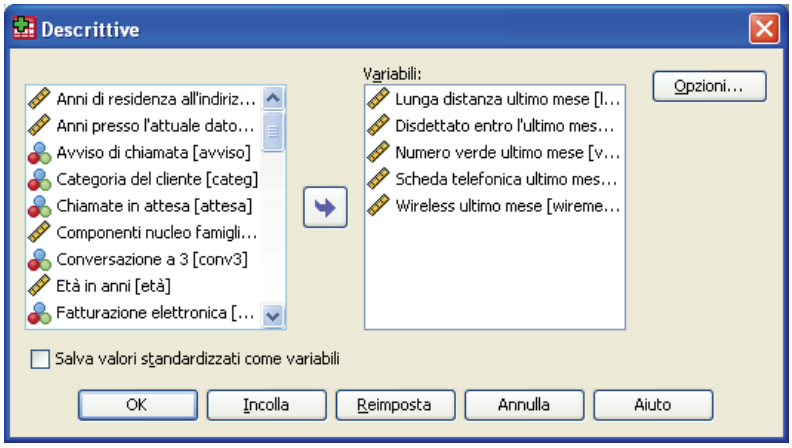
Assunzioni. La maggior parte delle statistiche disponibili (compresi i punteggi *z*) si fondano sulla teoria di normalità e possono essere utilizzate per le variabili quantitative (misurazioni a livello di intervallo o di rapporto) con distribuzioni simmetriche. Evitare variabili con categorie non ordinate o distribuzioni asimmetriche. La distribuzione dei punteggi *z* ha la stessa forma di quella dei dati originali. Pertanto, il calcolo dei punteggi *z* non rappresenta una soluzione per dati problematici.

Per ottenere statistiche descrittive

▶ Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Descrittive...

Figura 16-1
Finestra di dialogo Descrittive



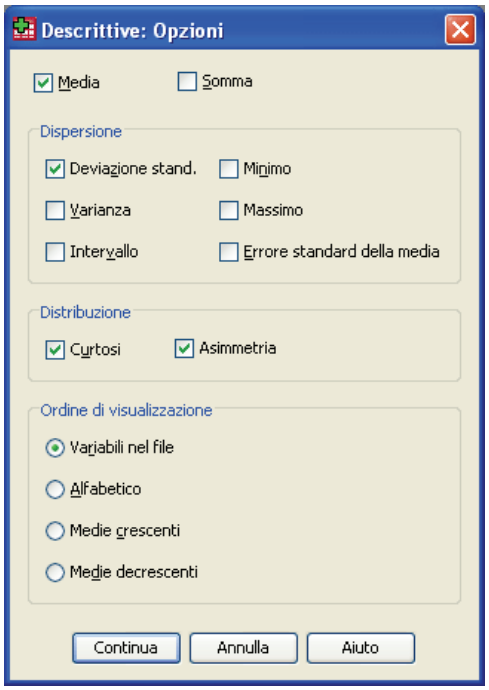
- Selezionare una o più variabili.

Se lo si desidera, è possibile:

- Selezionare Salva valori standardizzati come variabili per salvare i punteggi *z* come nuove variabili.
- Fare clic su Opzioni per ottenere le statistiche e l'ordine di visualizzazione facoltativi.

Descrittive: Opzioni

Figura 16-2
Finestra di dialogo Descrittive: Opzioni



Media e somma. Per impostazione predefinita, viene visualizzata la media o la media aritmetica.

Dispersione. Le statistiche che misurano la dispersione o la variazione dei dati includono deviazione standard, varianza, intervallo, valore minimo e massimo ed errore standard della media.

- **Deviazione stand..** La radice quadrata della varianza. La deviazione standard è una misura della dispersione intorno alla media espressa nella stessa unità di misura delle osservazioni. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, in una popolazione con distribuzione normale l'età media fosse 45 e la deviazione standard 10, il 95% dei casi cadrebbe fra 25 e 65 anni.
- **Varianza.** Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile.
- **Intervallo.** La differenza tra il valore massimo ed il valore minimo di una variabile numerica.
- **Minimo.** Il valore più basso assunto da una variabile numerica.
- **Massimo.** Il valore più alto di una variabile numerica.
- **Errore standard della media.** Una misura di quanto può variare il valore della media da campione a campione per campioni estratti dalla stessa distribuzione. Può essere utilizzata per confrontare genericamente la media osservata rispetto a un valore ipotizzato (ovvero, è possibile concludere che i due valori sono diversi se il rapporto della differenza rispetto all'errore standard è inferiore a -2 o maggiore di +2).

Distribuzione. Curtosi e asimmetria sono statistiche che caratterizzano la forma e la simmetria della distribuzione. Queste statistiche vengono visualizzate con i relativi errori standard.

- **Curtosi.** Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per la distribuzione normale, il valore della statistica di curtosi è zero. La curtosi positiva indica che le osservazioni sono più raggruppate e hanno una coda più lunga rispetto a quelle della distribuzione normale. La curtosi negativa indica che le osservazioni sono meno raggruppate e hanno code più corte.
- **Asimmetria.** Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con una notevole asimmetria positiva ha una lunga coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica lo scostamento dalla normale simmetria.

Ordine di visualizzazione. Per impostazione predefinita, le variabili vengono visualizzate nell'ordine in cui vengono selezionate. È inoltre possibile visualizzare le variabili in ordine alfabetico, per media crescente o per media decrescente.

Funzioni aggiuntive del comando DESCRIPTIVES

Il linguaggio della sintassi dei comandi consente inoltre di:

- Salvare i punteggi standardizzati (punteggi *z*) per alcune ma non per tutte le variabili (con il sottocomando `VARIABLES`).
- Specificare i nomi delle nuove variabili che contengono i punteggi standardizzati (con il sottocomando `VARIABLES`).

- Escludere dall'analisi i casi con valori mancanti per qualsiasi variabile (con il sottocomando `MISSING`).
- Ordinare le variabili visualizzate in base al valore di una statistica, non solo in base alla media (con il sottocomando `SORT`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
17

Esplora

La procedura Esplora produce statistiche riassuntive e visualizzazioni grafiche per tutti i casi o per singoli gruppi di casi. Risulta inoltre utile per numerose operazioni, ovvero screening dei dati, identificazione dei valori anomali, descrizione, verifica delle ipotesi e caratterizzazione delle differenze tra sottopopolazioni (gruppi di casi). Lo screening dei dati può evidenziare la presenza di valori insoliti, intervalli vuoti tra i dati o altri elementi specifici. L'esplorazione dei dati può consentire di determinare l'idoneità delle tecniche statistiche selezionate per l'analisi dei dati. L'esplorazione può evidenziare la necessità di eseguire una trasformazione dei dati se una particolare tecnica richiede una distribuzione normale. In alternativa è possibile utilizzare test non parametrici.

Esempio. Si consideri la distribuzione dei tempi in cui quattro gruppi di ratti imparano a uscire da un labirinto. Per ciascuno dei quattro gruppi, è possibile verificare se la distribuzione dei tempi è approssimativamente normale e se i quattro valori di varianza sono uguali. È inoltre possibile identificare i casi con i cinque tempi più lunghi e i cinque tempi più brevi. I grafici a scatole e i grafici ramo-foglia riassumono graficamente la distribuzione dei tempi di apprendimento per ciascun gruppo.

Statistiche e grafici. Media, mediana, media 5% trim, errore standard, varianza, deviazione standard, valore minimo e massimo, intervallo, distanza interquartilica, asimmetria e curtosi e i relativi errori standard, intervallo di confidenza per la media (e il livello di confidenza specificato), percentili, stimatore M di Huber, stimatore M di Andrew, stimatore M decrescente di Hampel, stimatore di Tukey a doppio peso, i cinque valori maggiori e i cinque valori minori, il test di Kolmogorov-Smirnov con il livello di significatività di Lilliefors per il test della normalità e il test di Shapiro-Wilk. Grafici a scatole, grafici ramo-foglia, istogrammi, grafici di normalità e grafici di variabilità contro intensità con test di Levene e trasformazioni.

Dati. La procedura Esplora può essere utilizzata per le variabili quantitative (livello di misurazione per intervallo o per rapporto). La variabile fattore, utilizzata per suddividere i dati in gruppi di casi, deve includere un numero ragionevole di valori distinti (categorie). Tali valori possono essere stringhe corte o numerici. La variabile etichetta di caso, utilizzata per etichettare i valori anomali in grafici a scatole, può essere una variabile stringa corta, stringa lunga (i primi 15 byte) o numerica.

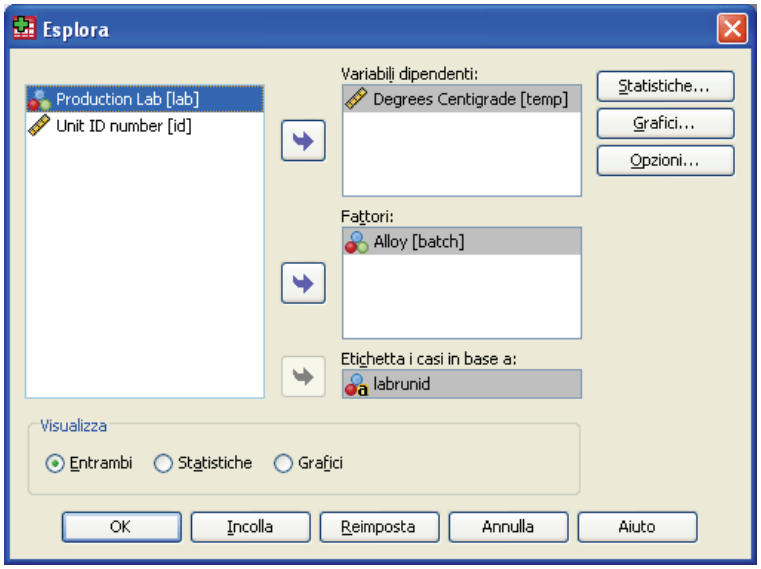
Assunzioni. La distribuzione dei dati non deve essere necessariamente simmetrica o normale.

Per esplorare i dati

▶ Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Esplora...

Figura 17-1
Finestra di dialogo Esplora



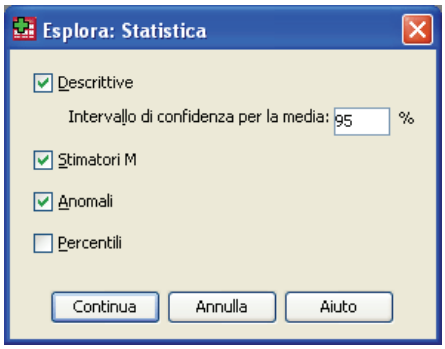
▶ Selezionare una o più variabili dipendenti.

Se lo si desidera, è possibile:

- Selezionare una o più variabili fattore i cui valori definiranno i gruppi di casi.
- Selezionare una variabile di identificazione per etichettare i casi.
- Fare clic su Statistiche per ottenere stimatori robusti, valori anomali, percentili e tabelle di frequenza.
- Fare clic su Grafici per ottenere istogrammi, grafici e test di probabilità normale e grafici di variabilità contro intensità con test di Levene.
- Fare clic su Opzioni per ottenere il trattamento dei valori mancanti.

Esplora: Statistica

Figura 17-2
Finestra di dialogo Esplora: Statistica



Descrittive. Queste misure di tendenza centrale e di dispersione vengono visualizzate per impostazione predefinita. Le misure di tendenza centrale indicano la posizione della distribuzione e includono la media, la mediana e la media 5% trim. Le misure di dispersione mostrano la dissimilarità dei valori e includono errore standard, varianza, deviazione standard, valore minimo e massimo, intervallo e distanza interquartilica. Le statistiche descrittive includono anche le misure della forma della distribuzione; l'asimmetria e la curtosi vengono visualizzate con i rispettivi errori standard. Viene visualizzato anche l'intervallo di confidenza al 95% per la media. È possibile specificare un diverso livello di confidenza.

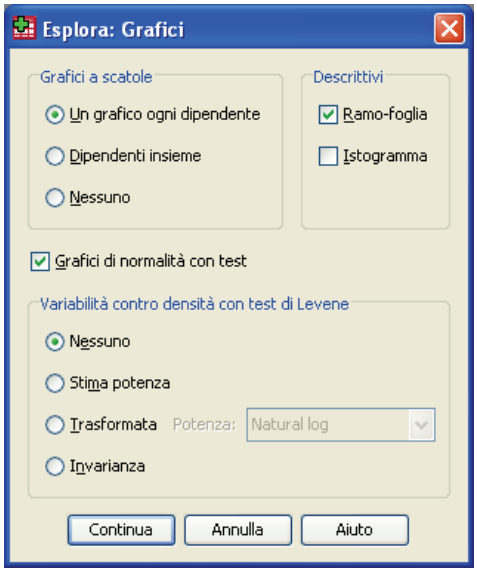
Stimatori M. Alternative valide alla media e alla mediana del campione per la valutazione della posizione. Gli stimatori calcolati differiscono per il peso applicato ai casi. Verranno visualizzati lo stimatore M di Huber, lo stimatore M di Andrews, lo stimatore M decrescente di Hampel e lo stimatore di Tukey a doppio peso.

Valori anomali. Consente di visualizzare i cinque valori maggiori e i cinque valori minori con le etichette dei casi.

Percentili. Consente di visualizzare i valori del 5°, 10°, 25°,50°, 75°, 90° e 95° percentile.

Esplora: Grafici

Figura 17-3
Finestra di dialogo Esplora: Grafici



Grafici a scatole. Queste alternative controllano la visualizzazione dei grafici a scatole quando sono presenti più variabili dipendenti. Un grafico ogni dipendente consente di generare una visualizzazione distinta per ciascuna variabile dipendente. All'interno della visualizzazione, vengono visualizzati grafici a scatole per ciascun gruppo definito da una variabile fattore. Dipendenti insieme consente di generare una visualizzazione distinta per ciascun gruppo definito da una variabile fattore. All'interno della visualizzazione compaiono grafici a scatole affiancati per ciascuna variabile dipendente. Questo tipo di grafico risulta particolarmente utile quando le singole variabili rappresentano una caratteristica misurata in tempi diversi.

Descrittive. Nel gruppo Descrittive è possibile scegliere grafici ramo-foglia e istogrammi.

Grafici di normalità con test. Consente di visualizzare grafici di probabilità normale e grafici di probabilità normale detrendizzati. Viene visualizzato il test di Kolmogorov-Smirnov con un livello di significatività di Lilliefors per il test della normalità. Se i pesi non interi sono specificati, la statistica di Shapiro-Wilk viene calcolata quando la dimensione campione pesata è compresa tra 3 e 50. Per pesi interi o non pesi, la statistica viene calcolata quando la dimensione del campione pesato è compresa tra 3 e 5.000.

Variabilità vs. intensità con test di Levene. Consente di controllare la trasformazione dei dati per i grafici di variabilità contro intensità. Per tutti i grafici di variabilità contro intensità vengono visualizzati l'inclinazione della curva di regressione e i test di Levene per l'omogeneità della varianza. Se si seleziona una trasformazione, i test di Levene si baseranno sui dati trasformati. Se non viene selezionata una variabile fattore, non verranno creati grafici di variabilità contro intensità. Stima potenza traccia i logaritmi naturali delle distanze interquartiliche verso i logaritmi naturali delle mediane di tutte le celle e inoltre una stima della potenza necessaria per trasformare i dati in modo da raggiungere varianze uguali in tutte le celle. Un grafico variabilità contro intensità consente di identificare la potenza di una trasformazione per stabilizzare (rendere maggiormente uguale) le varianze nei vari gruppi. Trasformata consente di selezionare un valore di potenza alternativo, seguendo o meno le indicazioni della stima di potenza, e di produrre i grafici dei dati trasformati. La distanza interquartilica e la media dei dati trasformati verranno tracciate in un grafico. Invarianza consente di ottenere grafici relativi ai dati semplici. Equivale a una trasformazione con potenza 1.

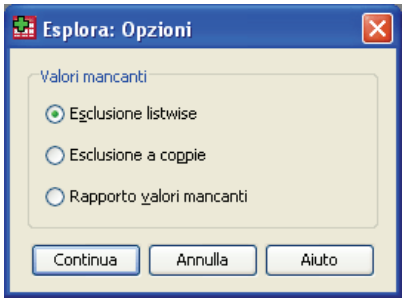
Esplora: potenza necessaria per la trasformazione dei dati

Si tratta delle trasformazioni di potenza per i grafici di variabilità contro intensità. Per trasformare i dati è necessario selezionare la potenza corrispondente. È possibile scegliere una delle seguenti opzioni:

- **Log naturale.** Trasformazione logaritmica naturale. È l'impostazione predefinita.
- **1/radice quadrata.** Per ciascun valore viene calcolato il reciproco della radice quadrata.
- **Reciproco.** Viene calcolato il reciproco di ciascun valore.
- **Radice quadrata.** Viene calcolata la radice quadrata di ciascun valore.
- **Quadrato.** Ciascun valore viene elevato al quadrato.
- **Cubo.** Ciascun valore viene elevato al cubo.

Esplora: Opzioni

Figura 17-4
Finestra di dialogo Esplora: Opzioni



Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile dipendente o fattore verranno esclusi da tutte le analisi. È l'impostazione predefinita.
- **Esclusione coppie.** I casi che non contengono valori mancanti per le variabili di un gruppo (cella) verranno inclusi nell'analisi per tale gruppo. Il caso può includere valori mancanti per le variabili utilizzate in altri gruppi.
- **Rapporto valori mancanti.** I valori mancanti per le variabili fattore vengono trattati come categoria distinta. Tutto l'output viene prodotto per questa categoria supplementare. Le tabelle di frequenza includono categorie per i valori mancanti. I valori mancanti per una variabile fattore vengono inclusi, ma etichettati come mancanti.

Funzioni aggiuntive del comando EXAMINE

La procedura Esplora usa la sintassi del comando `EXAMINE`. Il linguaggio della sintassi dei comandi consente inoltre di:

- Richiedere l'output totale e i grafici oltre all'output e ai grafici per i gruppi definiti dalle variabili di fattore (con il sottocomando `TOTAL`).
- Specificare una scala comune per un gruppo di grafici a scatole (con il sottocomando `SCALE`).
- Specificare le interazioni delle variabili di fattore (con il sottocomando `VARIABLES`).
- Specificare percentili diversi da quelli predefiniti (con il sottocomando `PERCENTILES`).
- Calcolare i percentili utilizzando uno dei cinque metodi (con il sottocomando `PERCENTILES`).
- Specificare una trasformazione di potenza per i grafici di variabilità vs. intensità (con il sottocomando `PLOT`).
- Specificare il numero di valori estremi da visualizzare (con il sottocomando `STATISTICS`).
- Specificare i parametri per i predittori M e i predittori robusti di posizione (con il sottocomando `MESTIMATORS`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
18

Tavole di contingenza

La procedura Tavole di contingenza consente di formare tabelle bivariate e a più dimensioni e fornisce una serie di test e misure di associazione per le tabelle bivariate. Il test o la misura da utilizzare vengono determinati in base alla struttura della tabella e al fatto che le categorie siano ordinate o meno.

Le statistiche e le misure delle tavole di contingenza vengono calcolate solo per le tabelle bivariate. Se si specifica una riga, una colonna o uno strato (variabile di controllo), verrà visualizzato un riquadro contenente le statistiche associate e le misurazioni per ciascun valore dello strato (o una combinazione di valori per due o più variabili di controllo). Ad esempio, se la variabile *sesso* è uno strato per la tabella della variabile *coniugato* (sì, no) rispetto alla variabile *tipo di vita* (ottima, soddisfacente, non soddisfacente), i risultati per la tabella bivariata per le donne vengono elaborati separatamente da quelli per gli uomini e quindi stampati come riquadri in successione.

Esempio. È possibile che i clienti rappresentati da piccole società siano più remunerativi per la vendita di servizi (per esempio addestramenti e consulenze) rispetto ai clienti rappresentati da società di grandi dimensioni? Mediante una tavola di contingenza è possibile scoprire che la maggior parte delle società di piccole dimensioni (con un numero di dipendenti inferiore a 500) fruttano alti profitti per i servizi, mentre la maggior parte delle grandi società (con oltre 2.500 dipendenti) fruttano profitti di scarsa entità.

Statistiche e misure di associazione. Chi-quadrato di Pearson, chi-quadrato del rapporto di verosimiglianza, test di associazione lineare-lineare, test esatto di Fisher, chi-quadrato corretto di Yates, *R* di Pearson, rho di Spearman, coefficiente di contingenza, phi, *V* di Cramér, lambda simmetrica e asimmetrica, tau di Goodman e Kruskal, coefficiente di incertezza, gamma, *D* di Somers, tau-*b* di Kendall, tau-*c* di Kendall, coefficiente eta, kappa di Cohen, stima del rischio relativo, rapporto odd, test di McNemar, statistiche di Cochran e Mantel-Haenszel.

Dati. Per definire le categorie di ciascuna variabile della tabella, utilizzare i valori di una variabile numerica o stringa (con una lunghezza massima di otto byte). Ad esempio, per la variabile *sesso*, è possibile codificare i dati come 1 e 2 oppure come *maschio* e *femmina*.

Assunzioni. Alcune statistiche e misure assumono categorie ordinate (dati ordinali) o valori quantitativi (dati misurati per intervallo o per rapporto), come indicato nella sezione sulle statistiche. Se le variabili della tabella prevedono categorie non ordinate (dati nominali), sono disponibili altri valori validi. Per le statistiche basate sul chi-quadrato (phi, *V* di Cramér e coefficiente di contingenza), i dati devono essere rappresentati da un campione casuale proveniente da una distribuzione multinomiale.

Nota: le variabili ordinali possono essere codici numerici che rappresentano categorie (ad esempio 1 = *basso*, 2 = *medio*, 3 = *alto*) oppure valori di stringa. Si suppone tuttavia che l'ordine alfabetico dei valori di stringa rifletta l'esatto ordine delle categorie. Ad esempio, per una variabile stringa

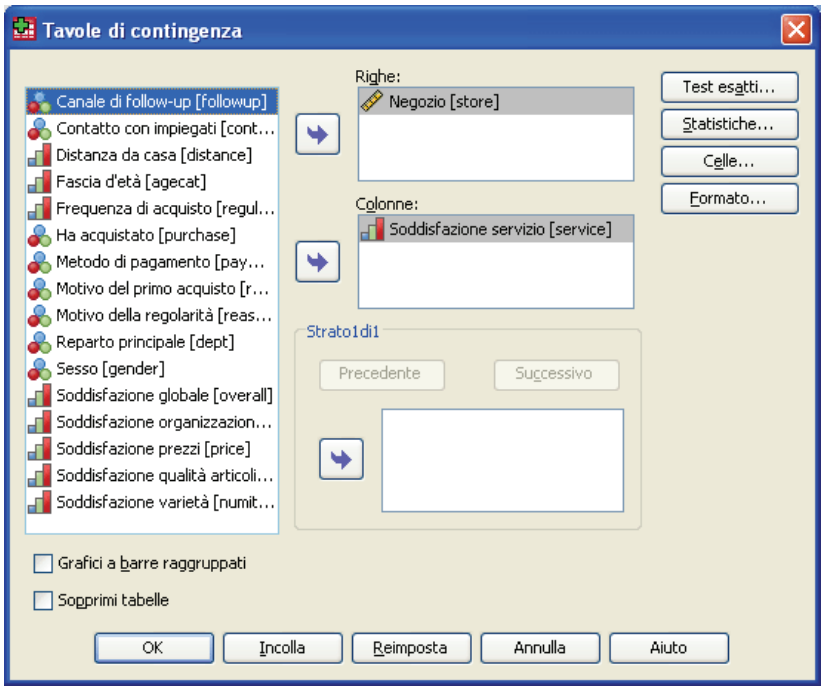
con i valori *basso*, *medio*, *alto*, l'ordine delle categorie viene interpretato come *alto*, *basso*, *medio*, ma questo non è l'ordine corretto. In generale, per rappresentare i dati ordinali, è più sicuro utilizzare i codici numerici.

Per ottenere tavole di contingenza

▶ Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Tavole di contingenza...

Figura 18-1
Finestra di dialogo Tavole di contingenza



▶ Selezionare una o più variabili di riga e una o più variabili di colonna.

Se lo si desidera, è possibile:

- Selezionare una o più variabili di controllo.
- Fare clic su Statistiche per ottenere test e misure di associazione per tabelle o sottotabelle bivariate.
- Fare clic su Celle per ottenere valori, percentuali e residui osservati e attesi.
- Fare clic su Formato per controllare l'ordine delle categorie.

Strati nelle tavole di contingenza

Se vengono selezionate una o più variabili di strato, verrà prodotta una tavola di contingenza distinta per ciascuna categoria di ciascuna variabile di strato (variabile di controllo). Ad esempio, se si dispone di una variabile di riga, una variabile di colonna e una variabile di strato con due

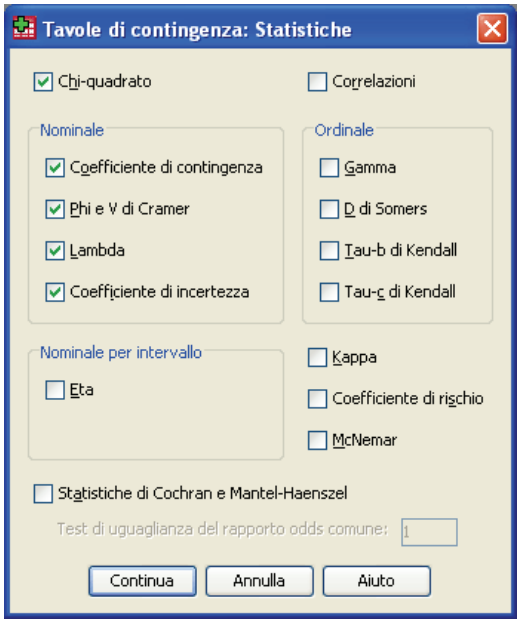
categorie, si otterrà una tabella bivariata per ciascuna categoria della variabile di strato. Per creare un altro strato di variabili di controllo, fare clic su Successivo. Verranno create sottotabelle per ogni combinazione delle categorie di ciascuna variabile del primo strato con ciascuna variabile del secondo e così via. Se sono richieste statistiche e misure di associazione, verranno applicate solo alle sottotabelle bivariate.

Tavole di contingenza: Grafici a barre raggruppati

Grafici a barre raggruppati. Nei grafici a barre raggruppati è possibile riepilogare i dati relativi a gruppi di casi. È disponibile un gruppo di barre per ciascun valore della variabile specificata in Righe. La variabile che definisce le barre contenute in ogni gruppo è quella specificata in Colonne. Per ciascun valore della variabile è disponibile una serie di barre con colori e motivi diversi. Se in Colonne o Righe si specificano più variabili, verrà prodotto un grafico a barre raggruppato per ciascuna combinazione delle due variabili.

Tavole di contingenza: Statistiche

Figura 18-2
Finestra di dialogo Tavole di contingenza: Statistiche



Chi-quadrato. Per tabelle con due righe e due colonne, scegliere Chi-quadrato per calcolare il chi-quadrato di Pearson, il chi-quadrato del rapporto di verosimiglianza, il test esatto di Fisher e il chi-quadrato corretto di Yates (correzione di continuità). Per le tabelle 2 × 2, il test esatto di Fisher viene calcolato quando una tabella non creata in base a righe o colonne mancanti in una tabella di dimensioni maggiori contiene una cella con una frequenza attesa minore di 5. Per tutte le altre tabelle 2 × 2 viene calcolato il chi-quadrato corretto di Yates. Per tabelle con un numero qualsiasi di righe e colonne, selezionare Chi-quadrato per calcolare il chi-quadrato di Pearson

e il chi-quadrato del rapporto di verosimiglianza. Se entrambe le variabili delle tabelle sono quantitative, l'opzione Chi-quadrato restituisce il test dell'associazione lineare.

Correlazioni. Per tabelle in cui sia le righe che le colonne contengono valori ordinati, l'opzione Correlazioni restituisce il coefficiente di correlazione di Spearman, rho (solo per dati numerici). Il coefficiente rho di Spearman è una misura di associazione tra punteggi di rango. Se entrambe le variabili delle tabelle (fattori) sono quantitative, Correlazioni restituisce il coefficiente di correlazione di Pearson, *r*, una misura dell'associazione lineare tra le variabili.

Nominale. Per i dati nominali (nessun ordine intrinseco, ad esempio cattolico, protestante, ebreo), è possibile selezionare il coefficiente Phi e V di Cramér, il Coefficiente di contingenza, Lambda (lambda simmetrico e asimmetrico e tau di Goodman e Kruskal), nonché il Coefficiente di incertezza.

- **Coefficiente di contingenza.** Una misura di associazione basata sul chi-quadrato. Questo coefficiente è sempre compreso tra 0 e 1, dove 0 indica nessuna associazione tra le variabili di riga e colonna e i valori vicini a 1 indicano un alto grado di associazione tra le variabili. Il valore massimo possibile dipende dal numero di righe e di colonne in una tabella.
- **Phi e V di Cramer.** Phi è una misura di associazione calcolata dividendo il chi-quadrato per la dimensione campionaria ed estraendo la radice quadrata del risultato. V di Cramér è una misura di associazione basata sul chi-quadrato.
- **Lambda.** Misura di associazione che riflette la riduzione proporzionale nell'errore quando i valori della variabile indipendente vengono usati per stimare quelli della variabile dipendente. Un valore pari a 1 significa che la variabile indipendente stima perfettamente la variabile dipendente. Un valore pari a 0 significa che la variabile indipendente non è di alcun aiuto nella stima della variabile dipendente.
- **Coefficiente di incertezza.** Misura di associazione che riflette la riduzione proporzionale nell'errore quando i valori di una variabile vengono usati per stimare i valori dell'altra. Un valore di 0,83, ad esempio, indica che la conoscenza di una variabile riduce dell'83% l'errore nella stima dei valori dell'altra variabile. La procedura calcola sia la versione simmetrica, sia quella asimmetrica.

Ordinale. Per tabelle in cui sia le righe che le colonne contengono valori ordinati, selezionare Gamma (gamma di ordine zero per tabelle a 2 vie e gamma condizionali per tabelle da 3 a 10 vie), Tau-b di Kendall e Tau-c di Kendall. Per desumere le categorie delle colonne delle righe, selezionare D di Somers.

- **Gamma.** Una misura di associazione simmetrica tra due variabili ordinali che varia tra -1 e 1. I valori prossimi al valore assoluto 1 indicano una forte relazione tra le due variabili. Valori prossimi allo zero indicano scarsità o assenza di relazione. In caso di tabelle a 2 vie verranno visualizzati gamma di ordine zero. Se una tavola di contingenza comprende più di due variabili, verrà calcolato un gamma condizionale per ciascuna sottotabella.
- **D di Somers.** Una misura di associazione tra due variabili ordinali. Varia fra -1 e 1, dove zero indica assenza di associazione e valori prossimi a 1 in valore assoluto indicano forte relazione. È una estensione asimmetrica di gamma dalla quale differisce solo per l'inclusione del numero di coppie non a pari merito nella variabile indipendente. La procedura calcola anche una versione simmetrica di questa statistica.
- **Tau-b di Kendall.** Una misura non parametrica di correlazione per variabili ordinali che tiene conto dei valori pari merito. Il segno del coefficiente indica la direzione della correlazione e il valore assoluto la sua intensità. Valori assoluti maggiori indicano correlazioni maggiori. I

valori possibili variano da -1 a +1, ma il valore -1 o +1 può solo essere ottenuto da tabelle quadrate.

- **Tau-c di Kendall.** Misura parametrica di correlazione per variabili ordinali che ignora i valori pari merito. Il segno del coefficiente indica la direzione della correlazione e il valore assoluto la sua intensità. Valori assoluti maggiori indicano correlazioni maggiori. I valori possibili variano da -1 a +1, ma il valore -1 o +1 può solo essere ottenuto da tabelle quadrate.

Nominale per intervallo. Se una variabile è categoriale e l'altra quantitativa, scegliere Eta. La variabile categoriale deve essere codificata numericamente.

- **Eta.** Una misura di associazione che varia fra 0 e 1, dove 0 indica assenza di associazione tra le variabili di riga e colonna e valori prossimi a 1 indicano un grado elevato di associazione. Eta è appropriata per una variabile dipendente misurata su una scala per intervallo e una variabile indipendente con numero limitato di categorie. Vengono calcolati due valori di Eta: il primo assume la variabile di riga, il secondo quella di colonna, come variabile misurata su una scala per intervallo.

Kappa. Il kappa di Cohen misura l'accordo tra due classificatori quando entrambi stanno classificando lo stesso oggetto. Un valore pari a 1 indica accordo perfetto. Un valore pari a 0 indica che l'accordo può essere considerato casuale. Il kappa è disponibile solo per le tabelle nelle quali entrambe le variabili usano valori di categoria identici e hanno lo stesso numero di categorie.

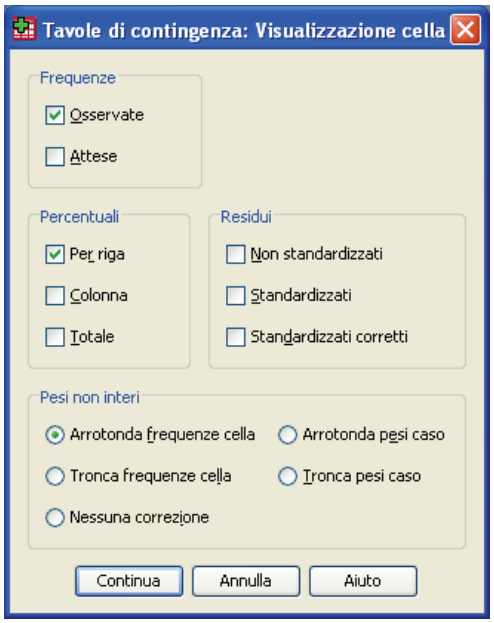
Rischio relativo. Nelle tabelle 2x2 il rischio relativo misura l'intensità dell'associazione tra la presenza di un fattore e la manifestazione di un evento. Un valore pari a 1 indica che il fattore non è associato all'evento. Il rapporto "odds ratio" può essere usato come stima del fattore di rischio quando la presenza del fattore è rara.

McNemar. Test non parametrico per due variabili dicotomiche correlate. Verifica la presenza di cambiamenti nelle risposte mediante la distribuzione chi-quadrato. Il test è molto utile in disegni sperimentali del tipo 'prima e dopo', per rilevare cambiamenti di risposta. Per tabelle quadrate di dimensioni maggiori, viene calcolato il test della simmetria di McNemar-Bowker.

Statistiche di Cochran e Mantel-Haenszel. Le statistiche di Cochran e Mantel-Haenszel possono essere usate come test di indipendenza fra un fattore binario e una variabile di risposta binaria. Le statistiche sono corrette per modelli covariati definiti da una o più variabili di controllo. Si noti che mentre altre statistiche vengono calcolate strato per strato, le statistiche di Cochran e Mantel-Haenszel vengono calcolate per tutti i livelli.

Tavole di contingenza: Visualizzazione cella

Figura 18-3
Finestra di dialogo Tavole di contingenza: Visualizzazione cella



Per facilitare l'individuazione di modelli di dati che danno origine a un test chi-quadrato significativo, la procedura per le tavole di contingenza visualizza le frequenze attese e tre tipi di residui (devianze) che misurano la differenza tra le frequenze osservate e quelle attese. Ogni cella della tabella può contenere qualsiasi combinazione dei conteggi, delle percentuali e dei residui selezionati.

Frequenze. Il numero di casi effettivamente osservati e il numero di casi attesi se le variabili di riga e di colonna sono reciprocamente indipendenti.

Percentuali. Le percentuali possono essere aggiunte nelle righe o nelle colonne. Sono disponibili anche le percentuali del numero totale di casi rappresentati nella tabella (a strato unico).

Residui. I residui semplici non standardizzati forniscono la differenza tra valori osservati e attesi. Sono inoltre disponibili residui standardizzati e standardizzati corretti.

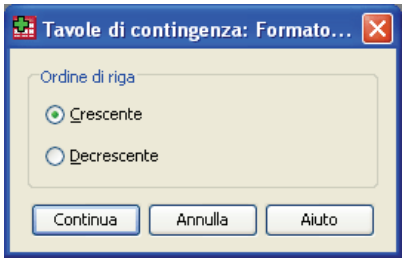
- **Non standardizzati.** La differenza fra un valore osservato e un valore previsto. Per valore atteso si intendo il numero di casi atteso nella cella in assenza di relazione tra le due variabili. Un residuo positivo indica che ci sono più casi nella cella di quanti ce ne sarebbero se le variabili di riga e di colonna fossero indipendenti.
- **Standardizzati.** Il residuo diviso per una stima della deviazione standard. Il residuo standardizzato, conosciuto anche come residuo di Pearson, ha media 0 e deviazione standard 1.
- **Standardizzati corretti.** Il residuo di una cella (valore osservato meno valore atteso) diviso per una stima del suo errore standard. Viene espresso in unità di deviazione standard sopra o sotto la media.

Pesi non interi. I conteggi di cella in genere sono valori interi, in quanto rappresentano il numero di casi in ogni cella. Se, tuttavia, il file di dati è attualmente ponderato in base a una variabile di ponderazione con valori frazionari (ad esempio, 1,25), i conteggi di cella possono essere espressi anche in valori frazionari. È possibile troncare o arrotondare i valori prima o dopo aver calcolato i conteggi di cella oppure utilizzare conteggi di cella frazionari per la visualizzazione delle tabelle e dei calcoli statistici.

- **Arrotonda frequenze cella.** I pesi dei casi vengono usati come tali, mentre i pesi accumulati nelle celle vengono arrotondati prima di calcolare qualsiasi statistica.
- **Tronca conteggi cella.** I pesi dei casi sono usati come tali, ma i pesi accumulati nelle celle sono troncati prima di calcolare qualunque statistica.
- **Arrotonda pesi caso.** I pesi dei casi vengono arrotondati prima dell'uso.
- **Tronca pesi caso.** I pesi dei casi sono troncati prima dell'uso.
- **Nessuna correzione.** I pesi di caso vengono utilizzati come sono e vengono utilizzati anche i conteggi di cella frazionari. Quando tuttavia è richiesto l'utilizzo dell'opzione Statistiche esatte (disponibile solo con l'opzione Test esatti), i pesi di caso delle celle verranno troncati o arrotondati prima del calcolo delle statistiche esatte.

Tavole di contingenza: Formato tabella

Figura 18-4
Finestra di dialogo Tavole di contingenza: Formato tabella



È possibile disporre le righe nell'ordine crescente o decrescente dei valori della variabile di riga.

Capitolo
19

Riassumi

La procedura Riassumi consente di calcolare le statistiche di sottogruppo per le variabili all'interno delle categorie di una o più variabili di raggruppamento. Tutti i livelli della variabile di raggruppamento vengono incrociati. È possibile scegliere l'ordine in cui vengono visualizzate le statistiche. Per ciascuna variabile di tutte le categorie verranno inoltre visualizzate le statistiche riassuntive. I valori di ciascuna categoria possono essere inseriti nell'elenco o eliminati, ma negli insiemi di dati di grandi dimensioni, è possibile scegliere di elencare solo i primi *n* casi.

Esempio. Qual è l'importo medio delle vendite per area e industria del cliente? Si potrebbe scoprire che l'importo medio delle vendite è leggermente superiore nell'area occidentale rispetto alle altre aree e che ai clienti di quest'area è associato l'importo medio più alto.

Statistiche. Somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del conteggio totale, percentuale della somma di gruppo, percentuale dei conteggi di gruppo, media geometrica, media armonica.

Dati. Le variabili di raggruppamento sono variabili categoriali che possono contenere valori stringa o numerici. Il numero di categorie dovrebbe essere limitato. Le altre variabili dovrebbero essere classificabili.

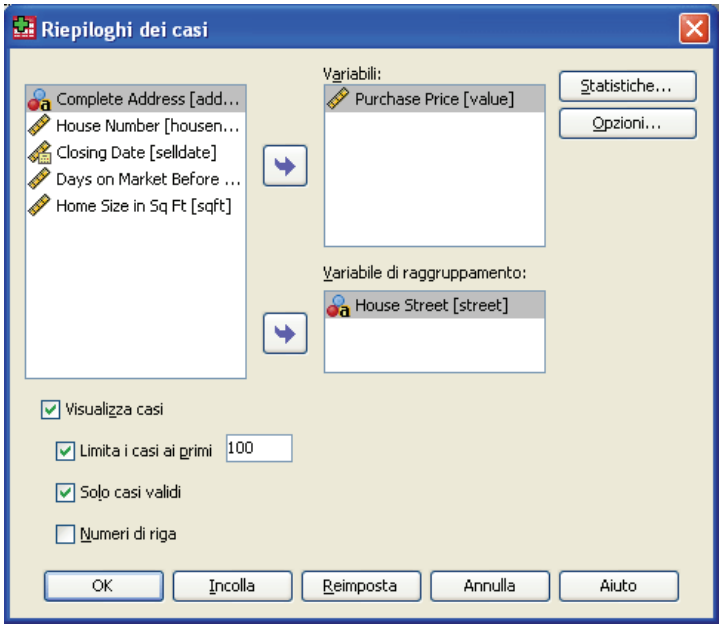
Assunzioni. Alcune delle statistiche di sottogruppo facoltative, quali la media e la deviazione standard, sono basate sulla teoria della normalità e sono idonee per le variabili quantitative con distribuzione simmetrica dei dati. La mediana e l'intervallo sono statistiche robuste, idonee per le variabili quantitative che possono o meno soddisfare l'assunzione di normalità.

Per ottenere riepiloghi dei casi

▶ Dai menu, scegliere:

Analizza
  Report
    Riepiloghi dei casi...

Figura 19-1
Finestra di dialogo Riepiloghi dei casi



► Selezionare una o più variabili.

Se lo si desidera, è possibile:

- Selezionare una o più variabili di raggruppamento per suddividere i dati in sottogruppi.
- Fare clic su Opzioni per modificare il titolo dell'output, aggiungere una didascalia al di sotto dell'output o escludere casi con valori mancanti.
- Fare clic su Statistiche per visualizzare statistiche facoltative.
- Selezionare Visualizza casi per visualizzare un elenco dei casi inclusi in ciascun sottogruppo. Per impostazione predefinita, vengono elencati solo i primi 100 casi nel file. È possibile aumentare o ridurre il valore di Limita i casi ai primi *n* oppure deselezionare l'opzione per visualizzare l'elenco di tutti i casi.

Riassumi: Opzioni

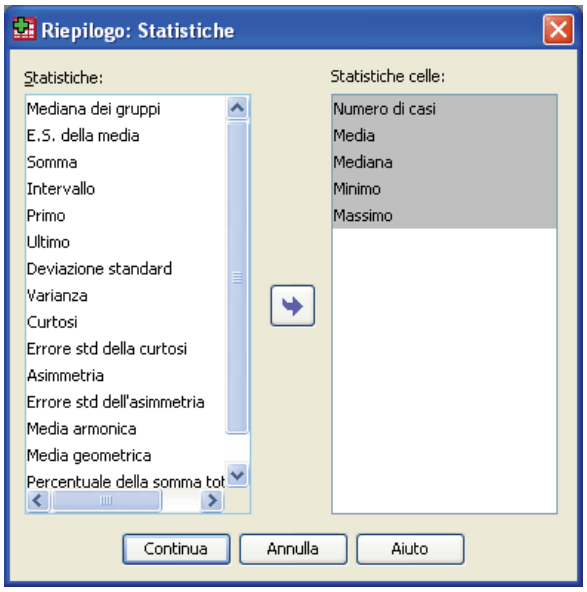
Figura 19-2
Finestra di dialogo Opzioni

SPSS consente di modificare il titolo dell'output o di aggiungere una didascalia che verrà visualizzata sotto alla tabella di output. È possibile controllare gli a capo automatici dei titoli e delle didascalie digitando \n dove si desidera inserire un'interruzione di riga nel testo.

È inoltre possibile scegliere di visualizzare o eliminare i sottototali e di includere o escludere casi con valori mancanti per qualsiasi variabile utilizzata nelle analisi. È spesso consigliabile contrassegnare nell'output i casi mancanti utilizzando un punto o un asterisco. Immettere un carattere, una frase o codice che si desidera venga visualizzato per indicare che un valore è mancante. In caso contrario, ai casi mancanti non verrà applicato alcun identificatore nell'output.

Riassumi: Statistiche

Figura 19-3
Finestra di dialogo delle statistiche dei report riassuntive



È possibile scegliere una o più delle seguenti statistiche di sottogruppo per le variabili all'interno di ogni categoria di ciascuna variabile di raggruppamento: somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del conteggio totale, percentuale della somma di gruppo, percentuale dei conteggi di gruppo, media geometrica, media armonica. L'ordine in cui compaiono le statistiche nell'elenco Statistiche di cella corrisponde all'ordine in cui verranno visualizzate nell'output. Per ciascuna variabile vengono visualizzate anche le statistiche riassuntive in tutte le categorie.

Primo. Visualizza il primo valore incontrato nel file di dati.

Media geometrica. La radice ennesima del prodotto dei valori, dove n è il numero di casi.

Mediana dei gruppi. La mediana calcolata su valori che rappresentano gruppi. Ad esempio la mediana della fascia di età.

Media armonica. Usata per stimare una dimensione media dei gruppi quando le dimensioni campionarie dei gruppi non sono uguali. La media armonica è il numero di campioni diviso per la somma dei reciproci delle dimensioni campionarie.

Curtosi. Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per la distribuzione normale, il valore della statistica di curtosi è zero. La curtosi positiva indica che le osservazioni sono più raggruppate e hanno una coda più lunga rispetto a quelle della distribuzione normale. La curtosi negativa indica che le osservazioni sono meno raggruppate e hanno code più corte.

Ultimo. Visualizza l'ultimo valore di dati incontrato nel file di dati.

Massimo. Il valore più alto di una variabile numerica.

Media. Una misura di tendenza centrale. La somma dei valori di tutte le osservazioni divisa per il numero di osservazioni. Viene anche detta media aritmetica.

Mediana. È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori eccezionalmente bassi o alti.

Minimo. Il valore più basso assunto da una variabile numerica.

N. Il numero di casi (osservazioni o record).

% del numero di casi totale. Percentuale del numero di casi totale in ogni categoria.

% della somma totale. Percentuale della somma totale in ogni categoria.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con una notevole asimmetria positiva ha una lunga coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica lo scostamento dalla normale simmetria.

Errore standard della curtosi. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore positivo elevato per la curtosi indica che le code della distribuzione sono più lunghe di quelle di una distribuzione normale; un valore negativo per la curtosi indica code più corte, simili a quelle di una distribuzione uniforme a forma di scatola.

Errore standard dell'asimmetria. Il rapporto fra l'asimmetria di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore positivo elevato per l'asimmetria indicata una coda a destra lunga; un valore negativo estremo indica una coda a sinistra lunga.

Somma. La somma o il totale di tutti i valori non mancanti di tutti i casi.

Varianza. Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile.

Capitolo
20

Medie

La procedura Medie consente di calcolare le medie dei sottogruppi e le statistiche univariate correlate per le variabili dipendenti all'interno delle categorie di una o più variabili indipendenti. È inoltre possibile ottenere analisi univariate della varianza, eta e test di linearità.

Esempio. Misurare la quantità media di grasso assorbita da tre diversi tipi di olii alimentari ed eseguire un'analisi univariata della varianza per verificare se le medie differiscono.

Statistiche. Somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del conteggio totale, percentuale della somma di gruppo, percentuale dei conteggi di gruppo, media geometrica, media armonica. Le opzioni includono analisi della varianza, eta, eta quadrato, e test di linearità R e R^2.

Dati. Le variabili dipendenti sono quantitative e le variabili indipendenti sono categoriali. I valori delle variabili categoriali possono essere di tipo numerico o stringa.

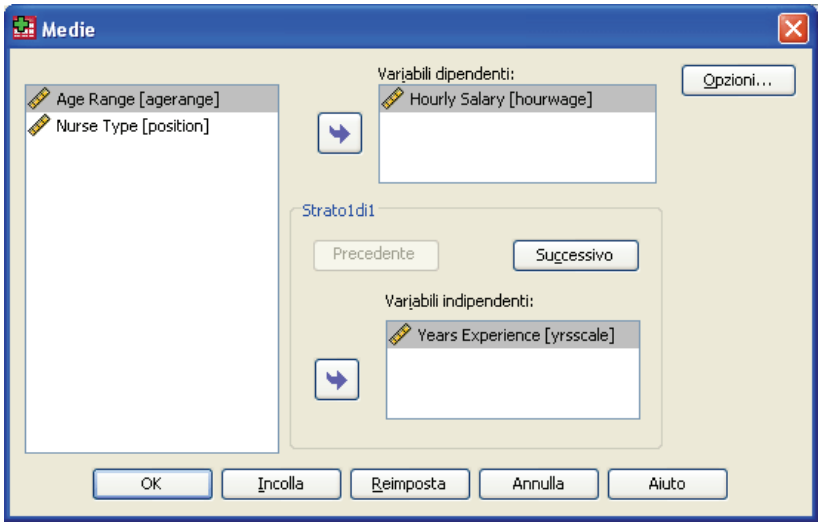
Assunzioni. Alcune delle statistiche di sottogruppo facoltative, quali la media e la deviazione standard, sono basate sulla teoria della normalità e sono idonee per le variabili quantitative con distribuzione simmetrica dei dati. La mediana è una statistica robusta, idonea per le variabili quantitative che possono o meno soddisfare l'ipotesi di normalità. L'analisi della varianza è robusta per quanto riguarda le alterazioni della normalità, ma i dati in ciascuna cella devono essere simmetrici. L'analisi della varianza assume inoltre che i gruppi provengano da popolazioni con valori di varianza uguali. Per verificare questa ipotesi, utilizzare il test di omogeneità della varianza di Levene, disponibile nella procedura ANOVA univariata.

Per ottenere le medie dei sottogruppi

- Dai menu, scegliere:

Analizza
  Confronta medie
    Medie...

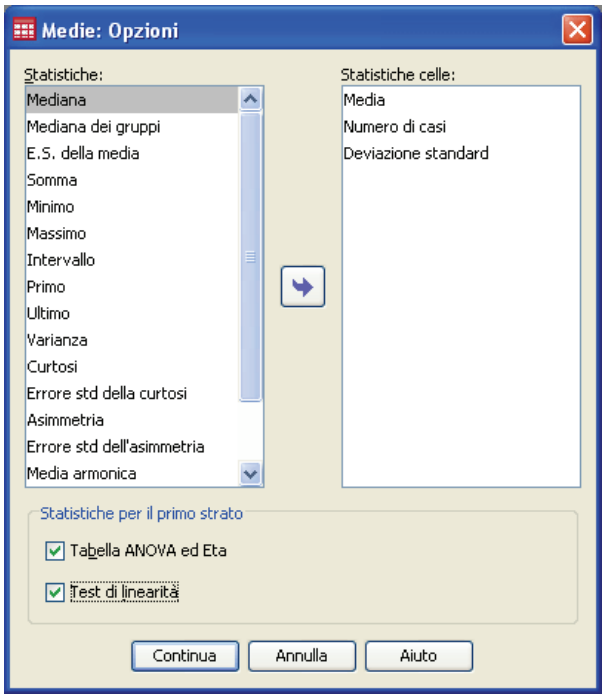
Figura 20-1
Finestra di dialogo Medie



- Selezionare una o più variabili dipendenti.
- Usare uno dei seguenti metodi per selezionare le variabili categoriali indipendenti:
  - Selezionare una o più variabili indipendenti. Per ciascuna variabile indipendente vengono visualizzati risultati distinti.
  - Selezionare uno o più strati di variabili indipendenti. Ogni strato suddivide ulteriormente il campione. Se è presente una sola variabile indipendente nello Strato 1 e una sola nello Strato 2, i risultati verranno visualizzati in una tabella incrociata e non in tabelle distinte per ciascuna variabile indipendente.
- Oppure fare clic su Opzioni per ottenere statistiche facoltative, analisi della tabella di varianza, età, età quadrato, R e R^2.

Medie: Opzioni

Figura 20-2
Finestra di dialogo Medie: Opzioni



È possibile scegliere una o più delle seguenti statistiche di sottogruppo per le variabili all'interno di ogni categoria di ciascuna variabile di raggruppamento: somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del numero*N* totale, percentuale della somma, percentuale del *numero di casi* in, media geometrica, media armonica. È possibile modificare l'ordine in cui compaiono le statistiche per i sottogruppi. L'ordine in cui compaiono le statistiche nella lista Statistiche di cella corrisponde all'ordine in cui verranno visualizzate nell'output. Per ciascuna variabile vengono visualizzate anche le statistiche riassuntive in tutte le categorie.

Primo. Visualizza il primo valore incontrato nel file di dati.

Media geometrica. La radice ennesima del prodotto dei valori, dove n è il numero di casi.

Mediana dei gruppi. La mediana calcolata su valori che rappresentano gruppi. Ad esempio la mediana della fascia di età.

Media armonica. Usata per stimare una dimensione media dei gruppi quando le dimensioni campionarie dei gruppi non sono uguali. La media armonica è il numero di campioni diviso per la somma dei reciproci delle dimensioni campionarie.

Curtosi. Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per la distribuzione normale, il valore della statistica di curtosi è zero. La curtosi positiva indica che le osservazioni sono più raggruppate e hanno una coda più lunga rispetto a quelle della distribuzione normale. La curtosi negativa indica che le osservazioni sono meno raggruppate e hanno code più corte.

Ultimo. Visualizza l'ultimo valore di dati incontrato nel file di dati.

Massimo. Il valore più alto di una variabile numerica.

Media. Una misura di tendenza centrale. La somma dei valori di tutte le osservazioni divisa per il numero di osservazioni. Viene anche detta media aritmetica.

Mediana. È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori eccezionalmente bassi o alti.

Minimo. Il valore più basso assunto da una variabile numerica.

N. Il numero di casi (osservazioni o record).

Percentuale del numero di casi totale. Percentuale del numero di casi totale in ogni categoria.

Percentuale della somma totale. Percentuale della somma totale in ogni categoria.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con una notevole asimmetria positiva ha una lunga coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica lo scostamento dalla normale simmetria.

Errore standard della curtosi. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore positivo elevato per la curtosi indica che le code della distribuzione sono più lunghe di quelle di una distribuzione normale; un valore negativo per la curtosi indica code più corte, simili a quelle di una distribuzione uniforme a forma di scatola.

Errore standard dell'asimmetria. Il rapporto fra l'asimmetria di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore positivo elevato per l'asimmetria indicata una coda a destra lunga; un valore negativo estremo indica una coda a sinistra lunga.

Somma. La somma o il totale di tutti i valori non mancanti di tutti i casi.

Varianza. Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile.

Statistiche per il primo strato

Tabella ANOVA ed Eta. Visualizza una tabella di analisi della varianza univariata e calcola le misure di associazione eta ed eta quadrato per ogni variabile indipendente del primo strato.

Test di linearità. Calcola la somma dei quadrati, i gradi di libertà e la media dei quadrati associata alle componenti lineari e non lineari, nonché il rapporto F, R e R-quadrato. La linearità non viene calcolata se la variabile indipendente è una stringa corta.

Capitolo
21

Cubi OLAP

La procedura Cubi OLAP (Online Analytical Processing) consente di calcolare i totali, le medie e le altre statistiche univariate per le variabili riassunte continue all'interno delle categorie di una o più variabili di raggruppamento categoriali. Nella tabella viene creato uno strato distinto per ciascuna categoria di ogni variabile di raggruppamento.

Esempio. Le vendite totali e medie di diverse aree e le linee di prodotti all'interno delle aree.

Statistiche. Somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del numero di casi totale, percentuale della somma totale entro variabili di raggruppamento, percentuale del numero di casi totale entro variabili di raggruppamento, media geometrica, media armonica.

Dati. Le variabili riassunte sono quantitative (variabili continue misurate su una scala di intervallo o di rapporto) e le variabili di raggruppamento sono categoriali. I valori delle variabili categoriali possono essere di tipo numerico o stringa.

Assunzioni. Alcune delle statistiche di sottogruppo facoltative, quali la media e la deviazione standard, sono basate sulla teoria della normalità e sono idonee per le variabili quantitative con distribuzione simmetrica dei dati. La mediana e l'intervallo sono statistiche robuste, idonee per le variabili quantitative che possono o meno soddisfare l'ipotesi di normalità.

Per ottenere cubi OLAP

- Dai menu, scegliere:

Analizza
  Report
    Cubi OLAP...

Figura 21-1
Finestra di dialogo Cubi OLAP

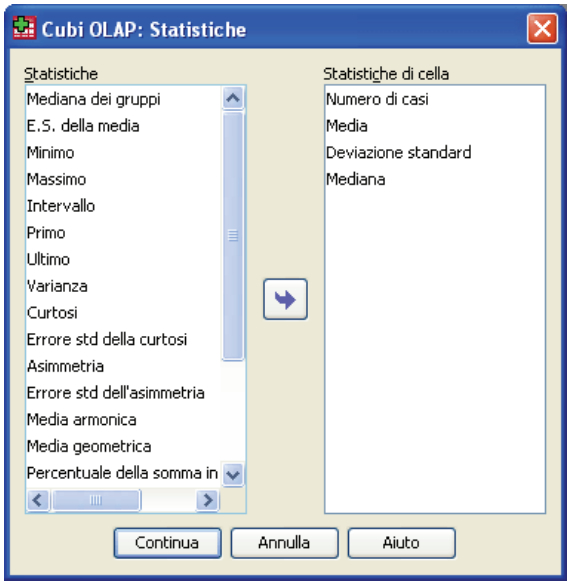
- Selezionare una o più variabili riassunte continue.
- Selezionare una o più variabili categoriali di raggruppamento

Oppure:

- Selezionare statistiche riassuntive diverse (fare clic su Statistiche). Prima di selezionare le statistiche riassuntive è necessario selezionare una o più variabili di raggruppamento.
- Calcolare differenze tra coppie di variabili e coppie di gruppi definiti da una variabile di raggruppamento (fare clic su Differenze).
- Creare titoli di tabella personalizzati (fare clic su Titolo).

Cubi OLAP: Statistiche

Figura 21-2
Finestra di dialogo Cubi OLAP: Statistiche



È possibile scegliere una o più delle seguenti statistiche di sottogruppo per le variabili riassunte all'interno di ogni categoria di ciascuna variabile di raggruppamento: Somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del numero di casi totale, percentuale della somma totale entro variabili di raggruppamento, percentuale del numero di casi totale entro variabili di raggruppamento, media geometrica, media armonica.

È possibile modificare l'ordine in cui compaiono le statistiche per i sottogruppi. L'ordine in cui compaiono le statistiche nella lista Statistiche di cella corrisponde all'ordine in cui verranno visualizzate nell'output. Per ciascuna variabile vengono visualizzate anche le statistiche riassuntive in tutte le categorie.

Primo. Visualizza il primo valore incontrato nel file di dati.

Media geometrica. La radice ennesima del prodotto dei valori, dove n è il numero di casi.

Mediana dei gruppi. La mediana calcolata su valori che rappresentano gruppi. Ad esempio la mediana della fascia di età.

Media armonica. Usata per stimare una dimensione media dei gruppi quando le dimensioni campionarie dei gruppi non sono uguali. La media armonica è il numero di campioni diviso per la somma dei reciproci delle dimensioni campionarie.

Curtosi. Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per la distribuzione normale, il valore della statistica di curtosi è zero. La curtosi positiva indica che le osservazioni sono più raggruppate e hanno una coda più lunga rispetto a quelle della distribuzione normale. La curtosi negativa indica che le osservazioni sono meno raggruppate e hanno code più corte.

Ultimo. Visualizza l'ultimo valore di dati incontrato nel file di dati.

Massimo. Il valore più alto di una variabile numerica.

Media. Una misura di tendenza centrale. La somma dei valori di tutte le osservazioni divisa per il numero di osservazioni. Viene anche detta media aritmetica.

Mediana. È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori eccezionalmente bassi o alti.

Minimo. Il valore più basso assunto da una variabile numerica.

N. Il numero di casi (osservazioni o record).

Percentuale del numero di casi in. Percentuale del numero di casi di una variabile di raggruppamento entro le categorie di altre variabili di raggruppamento. Se esiste una sola variabile di raggruppamento, questo valore è uguale alla percentuale del conteggio totale

% della somma in. Percentuale della somma di una variabile di raggruppamento entro le categorie di altre variabili di raggruppamento. Se esiste una sola variabile di raggruppamento, questo valore è uguale alla percentuale della somma totale.

% del numero di casi totale. Percentuale del numero di casi totale in ogni categoria.

% della somma totale. Percentuale della somma totale in ogni categoria.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con una notevole asimmetria positiva ha una lunga coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica lo scostamento dalla normale simmetria.

Errore standard della curtosi. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore positivo elevato per la curtosi indica che le code della distribuzione sono più lunghe di quelle di una distribuzione normale; un valore negativo per la curtosi indica code più corte, simili a quelle di una distribuzione uniforme a forma di scatola.

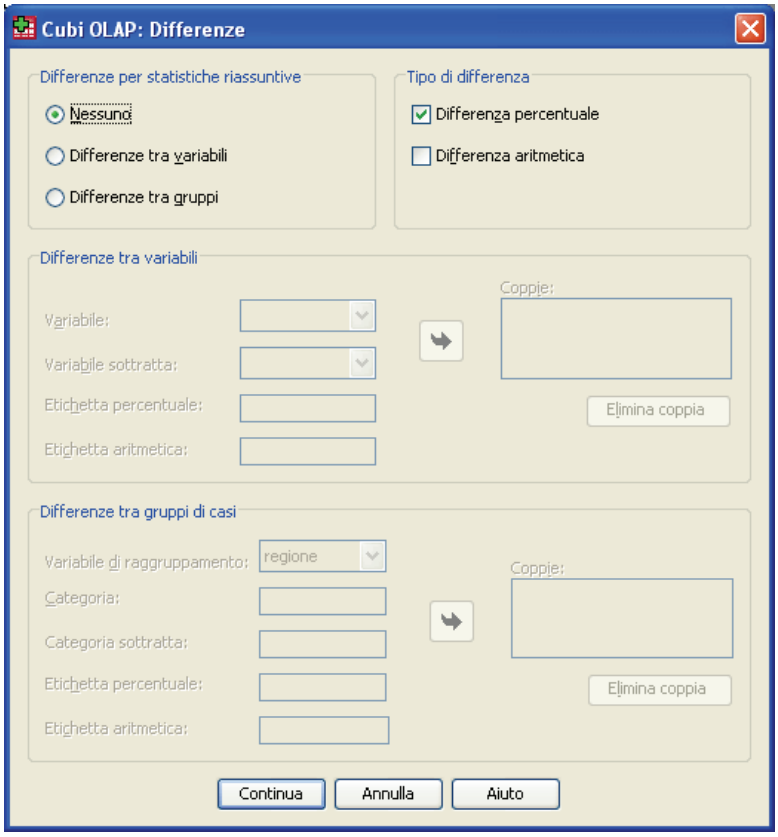
Errore standard dell'asimmetria. Il rapporto fra l'asimmetria di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore positivo elevato per l'asimmetria indicata una coda a destra lunga; un valore negativo estremo indica una coda a sinistra lunga.

Somma. La somma o il totale di tutti i valori non mancanti di tutti i casi.

Varianza. Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile.

Cubi OLAP: Differenze

Figura 21-3
Finestra di dialogo Cubi OLAP: Differenze



Questa finestra di dialogo consente di calcolare le differenze aritmetiche e percentuali tra variabili riassunte o tra gruppi definiti da una variabile di raggruppamento. Le differenze vengono calcolate per tutte le misure selezionate nella finestra di dialogo Cubi OLAP: Statistiche.

Differenze tra variabili. Consente di calcolare le differenze tra coppie di variabili. I valori delle statistiche riassuntive della seconda variabile di ogni coppia (la variabile sottratta) vengono sottratti dai valori delle statistiche riassuntive della prima variabile della coppia. Per le differenze percentuali, il valore della variabile riassunta della variabile sottratta viene utilizzato al denominatore. È necessario selezionare almeno due variabili riassunte nella finestra di dialogo principale prima di specificare le differenze tra variabili.

Differenze tra gruppi di casi. Consente di calcolare le differenze tra coppie di gruppi definiti da una variabile di raggruppamento. I valori delle statistiche riassuntive della seconda categoria di ogni coppia (la categoria sottratta) vengono sottratti dai valori delle statistiche riassuntive della prima categoria della coppia. Per le differenze percentuali, il valore delle statistiche riassuntive della categoria sottratta viene utilizzato al denominatore. È necessario selezionare una o più variabili di raggruppamento nella finestra di dialogo principale prima di specificare le differenze tra gruppi.

Cubi OLAP: Titolo

Figura 21-4
Finestra di dialogo Cubi OLAP: Titolo

È possibile modificare il titolo dell'output o aggiungere una didascalia che verrà visualizzata sotto la tabella dell'output. È inoltre possibile impostare gli a capo automatici dei titoli e delle didascalie digitando \n dove si desidera inserire un'interruzione di riga nel testo.

Capitolo
22

Test T

Sono disponibili tre tipi di test *t*:

Test T per campioni indipendenti (test T per due campioni). Consente di confrontare le medie di una variabile per due gruppi di casi. Vengono fornite statistiche descrittive per ciascun gruppo, il test di Levene di uguaglianza delle varianze, i valori *t* di uguaglianza e non uguaglianza della varianza e un intervallo di confidenza al 95% per la differenza tra le medie.

Test T per campioni appaiati (test T dipendente). Consente di confrontare le medie di due variabili per un singolo gruppo. Questo test viene utilizzato anche per disegni relativi a studi di confronti tra coppie o di casi di controllo. Vengono fornite statistiche descrittive per le variabili oggetto del test, la correlazione tra di esse, le statistiche descrittive per le differenze appaiate, il test *t* e un intervallo di confidenza al 95%.

Test T per un campione. Consente di confrontare la media di una variabile con un valore noto o un valore ipotizzato. Con il test *t* vengono visualizzate anche le statistiche descrittive per le variabili oggetto del test. L'output predefinito include un intervallo di confidenza al 95% per la differenza tra la media della variabile oggetto del test e il valore ipotizzato per il test.

T per campioni indipendenti

Il test T per campioni indipendenti consente di confrontare le medie relative a due gruppi di casi. Nel test, i soggetti dovrebbero essere assegnati in modo casuale a due gruppi. In questo modo, le eventuali differenze nella riposta saranno dovute alla modalità di elaborazione (o alla mancata elaborazione) e non ad altri fattori. Ciò non si verifica se si esegue il confronto tra il reddito medio di soggetti maschili e femminili. Non è infatti possibile assegnare in modo casuale una persona al sesso maschile o femminile. In questi casi, è necessario assicurarsi che le differenze relative ad altri fattori non comportino un mascheramento o l'incremento di differenze significative nelle medie. Le differenze nel reddito medio possono essere influenzate da fattori quali il livello di educazione e non solo dal sesso al quale appartengono i soggetti.

Esempio. I pazienti con pressione sanguigna alta vengono assegnati in modo casuale a un gruppo di controllo e a un gruppo di trattamento. Ai soggetti del gruppo di controllo vengono somministrate medicine innocue e ai soggetti del gruppo trattato viene somministrato un nuovo farmaco che si ritiene possa far diminuire la pressione sanguigna. Al termine di un trattamento di due mesi, viene utilizzato il test *t* per due campioni allo scopo di confrontare i valori medi della pressione sanguigna nel gruppo di controllo e nel gruppo trattato. La pressione di ogni paziente viene misurata una volta e ciascun paziente appartiene a un solo gruppo.

Statistiche. Per ogni variabile: dimensione campione, media, deviazione standard ed errore standard della media. Per le differenze fra le medie: media, errore standard e intervallo di confidenza (è possibile specificare il livello di confidenza). Test: test di Levene di uguaglianza delle varianze e test *t* di uguaglianza delle medie per la varianza comune e la varianza separata.

Dati. I valori della variabile quantitativa desiderata si trovano in una singola colonna del file di dati. Viene utilizzata una variabile di raggruppamento che include due valori per suddividere i casi in due gruppi. La variabile di raggruppamento può essere numerica (valori quali 1 e 2, o 6,25 e 12,5) oppure una stringa breve (ad esempio *sì* e *no*). In alternativa, è possibile utilizzare una variabile quantitativa, ad esempio *età*, per suddividere i casi in due gruppi specificando un punto di divisione (il punto di divisione 21 suddivide la variabile *età* in un gruppo con meno di 21 anni e in un gruppo con più di 21 anni).

Assunzioni. Per il test *t* di uguaglianza della varianza, le osservazioni dovrebbero essere rappresentate da campioni indipendenti e casuali derivati da distribuzioni normali con la stessa varianza di popolazione. Per il test *t* di inuguaglianza della varianza, le osservazioni dovrebbero essere campioni indipendenti e casuali derivati da distribuzioni normali. Il test *t* per due campioni è sufficientemente robusto per le deviazioni dalla normalità. Durante la verifica grafica delle distribuzioni, controllare che siano simmetriche e che non siano presenti valori anomali.

Per ottenere un test T per campioni indipendenti

▶ Dai menu, scegliere:

Analizza
  Confronta medie
    Test T: campioni indipendenti...

Figura 22-1
Finestra di dialogo Test T per campioni indipendenti

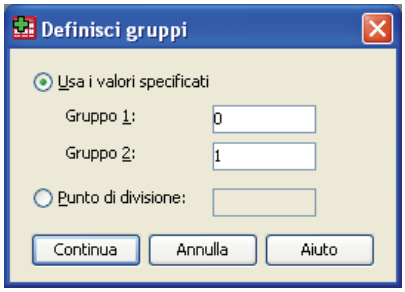
▶ Selezionare una o più variabili quantitative oggetto del test. Per ciascuna variabile viene calcolato un test *t* distinto.

▶ Selezionare una variabile di raggruppamento singola e fare clic su Definisci gruppi per specificare due codici per i gruppi che si desidera confrontare.

▶ Se necessario, fare clic su Opzioni per verificare in quale modo vengono considerati i dati mancanti e il livello dell'intervallo di confidenza.

Test T per campioni indipendenti: Definisci gruppi

Figura 22-2
Finestra di dialogo Definisci gruppi per le variabili numeriche



Per le variabili di raggruppamento numeriche, definire i due gruppi per il test *t* specificando due valori o un punto di divisione:

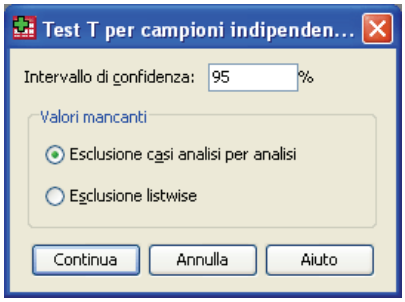
- **Usa i valori specificati.** Inserire un valore per Gruppo 1 e un altro valore per Gruppo 2. I casi con qualsiasi altro valore verranno esclusi dall'analisi. Non è necessario specificare numeri interi (ad esempio, 6,25 e 12,5 sono validi).
- **Punto di divisione.** Immettere un numero che suddivide i valori della variabile di raggruppamento in due insiemi Assegna i casi con valori minori o uguali a quello specificato a un gruppo e i casi con valori maggiori all'altro.

Figura 22-3
Finestra di dialogo Definisci gruppi per le variabili stringa

Per le variabili stringa di raggruppamento, inserire una stringa per Gruppo 1 e un'altra stringa per Gruppo 2, ad esempio *sì* e *no*. I casi che includono altre stringhe verranno esclusi dall'analisi.

Test T per campioni indipendenti: Opzioni

Figura 22-4
Finestra di dialogo Test T per campioni indipendenti: Opzioni



Intervallo di confidenza. Per impostazione predefinita, viene visualizzato un intervallo di confidenza del 95% per la differenza fra le medie. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Valori mancanti. Se durante un test su più variabili si riscontra in alcune di esse la presenza di dati mancanti, è possibile indicare alla procedura i casi da includere (o da escludere):

- **Esclusione casi analisi per analisi.** Per ciascun test *t* vengono utilizzati tutti i casi con dati validi per le variabili verificate. Le dimensioni del campione possono variare in base al test.
- **Esclusione listwise.** Per ciascun test *t* vengono utilizzati solo i casi con dati validi per tutte le variabili prese in considerazione nei test *t*. La dimensione del campione è costante nei vari test.

T per campioni appaiati

La procedura Test T per campioni appaiati consente di confrontare le medie di due variabili per un singolo gruppo. La procedura calcola le differenze tra i valori delle due variabili per ciascun caso e viene verificato se la media è diversa da 0.

Esempio. In uno studio su pazienti con valori elevati della pressione sanguigna, a tutti i pazienti è stata misurata la pressione all'inizio dello studio, è stato somministrato un trattamento e quindi la misurazione è stata ripetuta. Per ciascun soggetto sono quindi disponibili due misurazioni, in genere denominate *precedente* e *successiva*. Questo test viene utilizzato anche per disegni relativi a studi di confronti tra coppie o di casi di controllo, in cui ciascun record del file di dati contiene la risposta per il paziente e quella del soggetto di controllo corrispondente. In uno studio sulla pressione sanguigna, è necessario che l'età dei pazienti trattati corrisponda a quella dei controlli (a un paziente di 75 anni deve corrispondere un membro del gruppo di controllo di 75 anni).

Statistiche. Per ogni variabile: media, dimensione campione, deviazione standard ed errore standard della media. Per ciascuna coppia di variabili: correlazione, differenza media tra le medie, test *t* e intervallo di confidenza per la differenza tra medie (è possibile specificare il livello di confidenza). Deviazione standard ed errore standard della differenza fra medie.

Dati. Per ciascun test appaiato, specificare due variabili quantitative (livello di misura in base a intervallo o a rapporto). In uno studio di confronti tra coppie o di casi di controllo, la risposta per ciascun soggetto del test e per il soggetto di controllo corrispondente deve trovarsi nello stesso caso all'interno del file di dati.

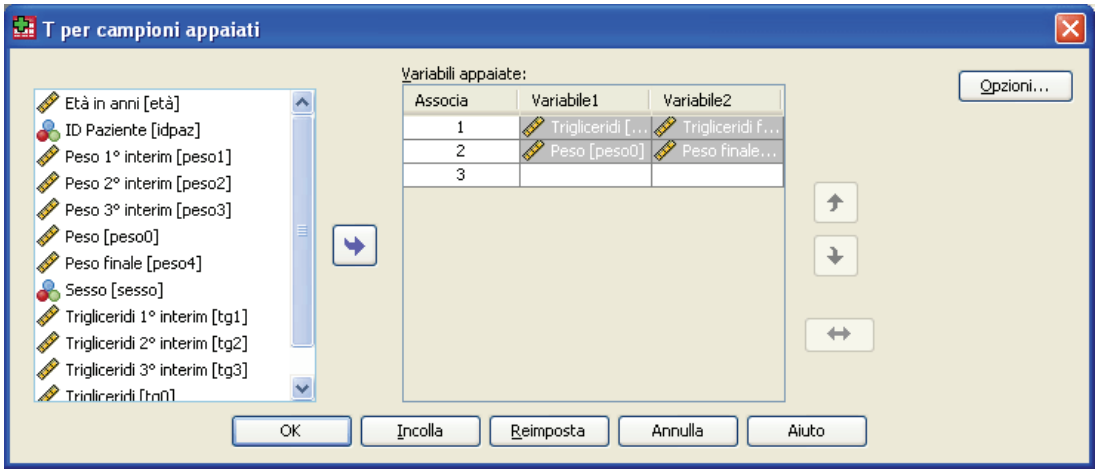
Assunzioni. Le osservazioni per ciascuna coppia devono essere effettuate nelle medesime condizioni. Le differenze tra medie devono essere distribuite normalmente. Le varianze di ciascuna variabile possono essere uguali o non uguali.

Per ottenere un test T per campioni appaiati

▶ Dai menu, scegliere:

Analizza
  Confronta medie
    Test T: campioni appaiati...

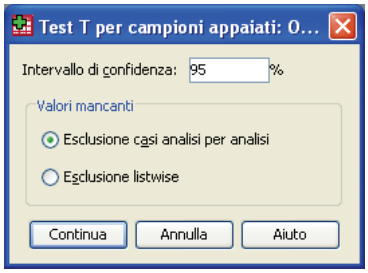
Figura 22-5
Finestra di dialogo Test T per campioni appaiati



- Selezionare una o più coppie di variabili
- Se necessario, fare clic su Opzioni per verificare in quale modo vengono considerati i dati mancanti e il livello dell'intervallo di confidenza.

Test T per campioni appaiati: Opzioni

Figura 22-6
Finestra di dialogo Test T per campioni appaiati: Opzioni



Intervallo di confidenza. Per impostazione predefinita, viene visualizzato un intervallo di confidenza del 95% per la differenza fra le medie. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Valori mancanti. Se durante un test su più variabili si riscontra in alcune di esse la presenza di dati mancanti, è possibile indicare alla procedura i casi da includere (o da escludere):

- **Esclusione casi analisi per analisi.** Per ciascun test *t* vengono utilizzati tutti i casi con dati validi per la coppia di variabili verificate. Le dimensioni del campione possono variare in base al test.
- **Esclusione listwise.** Per ciascun test *t* vengono utilizzati solo i casi che includono dati validi per tutte le coppie di variabili verificate. La dimensione del campione è costante nei vari test.

Test T per un campione

La procedura Test T per un campione consente di verificare se la media di una singola variabile è diversa da una costante specificata.

Esempi. Un ricercatore può verificare se il quoziente d'intelligenza (QI) medio di un gruppo di studenti è diverso da 100. Oppure, un produttore di cereali può prelevare un campione di scatole dalla linea di produzione e verificare se il peso medio dei campioni è diverso da 0,7 kg con un livello di confidenza al 95%.

Statistiche. Per ciascuna variabile oggetto del test: media, deviazione standard ed errore standard della media. La differenza media fra ciascun valore e il valore oggetto del test ipotizzato, un test *t* che verifica che la differenza sia uguale a 0 e un intervallo di confidenza per la differenza (è possibile specificare il livello di confidenza).

Dati. Per confrontare i valori di una variabile quantitativa con un valore oggetto del test ipotizzato, scegliere una variabile quantitativa e immettere un valore oggetto del test ipotizzato.

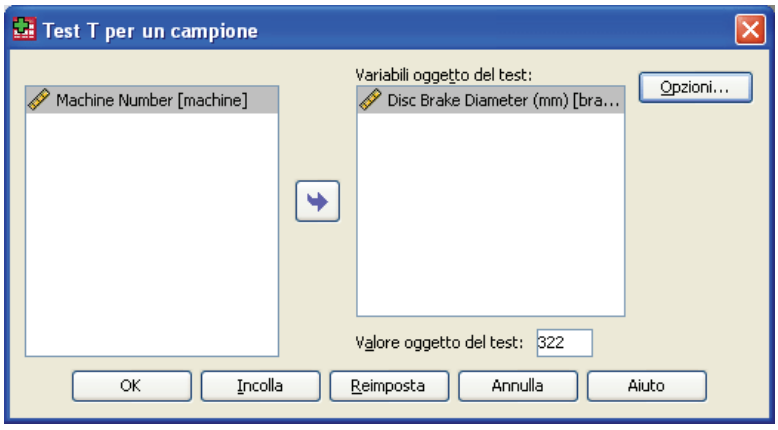
Assunzioni. In questo test si presume che i dati siano distribuiti in modo normale. Il test è tuttavia sufficientemente robusto per le deviazioni dalla normalità.

Per ottenere un test T per un campione

- Dai menu, scegliere:

Analizza
  Confronta medie
    Test T: campione unico...

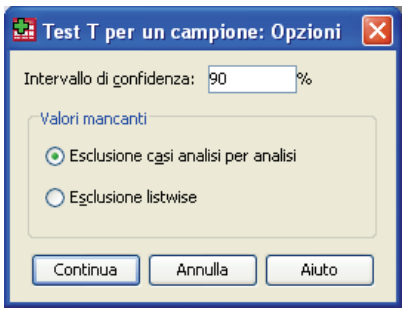
Figura 22-7
Finestra di dialogo Test T per un campione



- Selezionare una o più variabili da confrontare con lo stesso valore ipotizzato.
- Immettere un valore oggetto del test numerico rispetto al quale viene confrontata ciascuna media del campione.
- Se necessario, fare clic su Opzioni per verificare in quale modo vengono considerati i dati mancanti e il livello dell'intervallo di confidenza.

Test T per un campione: Opzioni

Figura 22-8
Finestra di dialogo Test T per un campione: Opzioni



Intervallo di confidenza. Per impostazione predefinita, viene visualizzato un intervallo di confidenza del 95% per la differenza fra la media e il valore oggetto del test ipotizzato. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Valori mancanti. Se durante un test su più variabili si riscontra in alcune di esse la presenza di dati mancanti, è possibile indicare alla procedura i casi da includere (o da escludere):

- **Esclusione casi analisi per analisi.** Per ciascun test *t* vengono utilizzati tutti i casi con dati validi per le variabili verificate. Le dimensioni del campione possono variare in base al test.
- **Esclusione listwise.** Per ciascun tes*t* vengono utilizzati solo casi con dati validi per tutte le variabili prese in considerazione nei *t* test richiesti. La dimensione del campione è costante nei vari test.

Opzioni aggiuntive del comando T-TEST

Il linguaggio della sintassi dei comandi consente inoltre di:

- Effettuare test di T per un campione e per campioni indipendenti tramite un unico comando.
- Confrontare ciascuna variabile con le variabili dell'elenco in test accoppiati (con il sottocomando `PAIRS`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
23

ANOVA univariata

La procedura ANOVA univariata produce un'analisi della varianza univariata per una variabile dipendente quantitativa in base a una singola variabile fattore (indipendente). L'analisi della varianza consente di verificare l'ipotesi di uguaglianza di più medie. Questa tecnica è un'estensione del test *t* per due campioni.

Oltre a determinare le differenze tra le medie, è possibile individuare la media che differisce dalle altre. Esistono due tipi di test per il confronto tra le medie: contrasti a priori e test post hoc. I contrasti sono test impostati *prima* di eseguire l'esperimento, mentre i test post hoc vengono effettuati *dopo* l'esecuzione dell'esperimento. È inoltre possibile verificare i trend presenti tra le categorie.

Esempio. Le ciambelle assorbono quantità variabili di grassi a seconda della modalità di cottura. È stato condotto un esperimento che prevede l'utilizzo di tre tipi di grassi: olio di semi di arachide, olio di mais e strutto. L'olio di semi di arachide e l'olio di mais sono grassi insaturi, mentre lo strutto è un grasso saturo. Oltre a determinare se la quantità di grassi assorbita dipende dal tipo di grasso utilizzato, è possibile impostare un contrasto a priori per determinare se la quantità di grassi assorbita differisce per i grassi saturi e insaturi.

Statistiche. Per ciascun gruppo: numero di casi, media, deviazione standard, errore standard della media, valore minimo e massimo e intervallo di confidenza al 95% per la media. Test di omogeneità della varianza di Levene, tabella di analisi della varianza e test robusti dell'uguaglianza delle medie per ciascuna variabile dipendente, contrasti a priori definiti dall'utente e test di intervallo post hoc e confronti multipli: Bonferroni, Sidak, differenze significative di Tukey, GT2 di Hochberg, Gabriel, Dunnett, procedura di Ryan-Einot-Gabriel-Welsch basata su test *F*, (R-E-G-W *F*), test a intervalli di Ryan-Einot-Gabriel-Welsch (R-E-G-W *Q*), T2 di Tamhane, T3 di Dunnett, Games-Howell, *C* di Dunnett, test a intervalli multipli di Duncan, Student-Newman-Keuls (S-N-K), *b* di Tukey, Waller-Duncan, Scheffée differenza meno significativa.

Dati. I valori delle variabili fattore devono essere interi e la variabile dipendente deve essere quantitativa (livello di misura per intervallo).

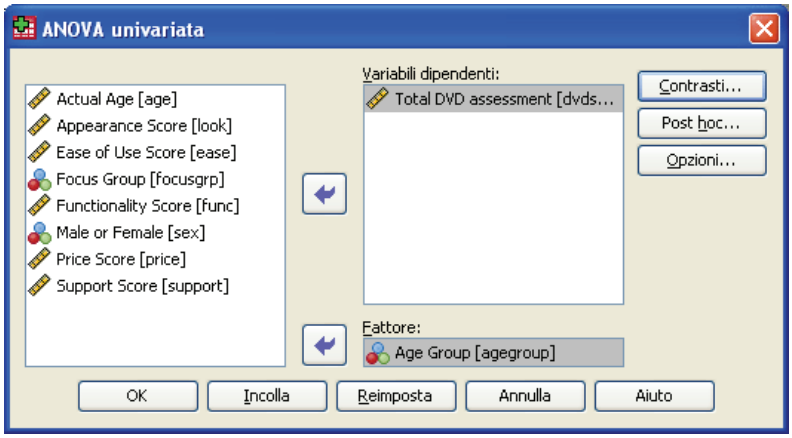
Assunzioni. Ciascun gruppo è un campione casuale indipendente prelevato da una popolazione normale. L'analisi della varianza è uno stimatore robusto degli scostamenti dalla normalità, anche se i dati devono essere simmetrici. I gruppi devono provenire da popolazioni con varianze uguali. Per verificare questa ipotesi, utilizzare il test dell'omogeneità della varianza di Levene.

Per ottenere un'analisi della varianza univariata

▶ Dai menu, scegliere:

Analizza
  Confronta medie
    ANOVA univariata...

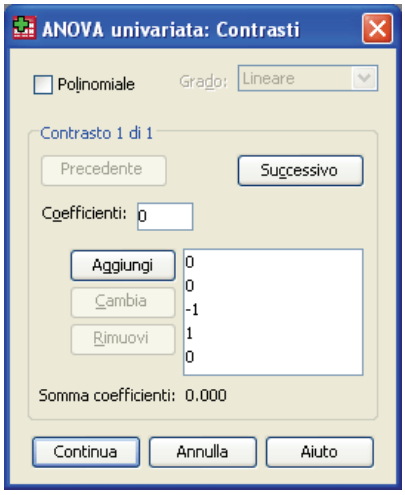
Figura 23-1
Finestra di dialogo ANOVA univariata



- Selezionare una o più variabili dipendenti.
- Selezionare una singola variabile fattore indipendente.

ANOVA univariata: Contrasti

Figura 23-2
Finestra di dialogo ANOVA univariata: Contrasti



È possibile suddividere le somme dei quadrati fra gruppi in componenti di trend oppure specificare contrasti a priori.

Polinomiale. Consente di suddividere le somme dei quadrati tra gruppi in componenti di trend. È possibile verificare un trend della variabile dipendente in tutti i livelli ordinati della variabile fattore. Ad esempio, è possibile verificare un trend lineare (crescente o decrescente) nei salari in tutti i livelli ordinati del grado di salario più elevato.

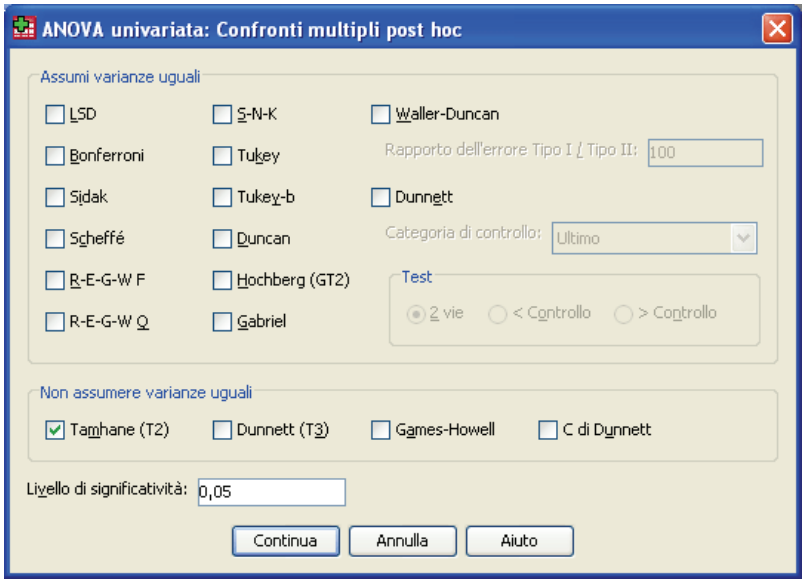
- **Grado.** È possibile scegliere un termine polinomiale di ordine 1, 2, 3, 4 o 5.

Coefficienti. Contrasti a priori definiti dall'utente da verificare con la statistica *t*. Specificare un coefficiente per ciascun gruppo (categoria) della variabile fattore e quindi fare clic su Aggiungi dopo aver inserito ciascuna voce. I nuovi valori verranno aggiunti alla fine dell'elenco dei coefficienti. Per specificare altri insiemi di contrasti, fare clic su Successivo. Utilizzare Successivo e Precedente per spostarsi tra gli insiemi di contrasti.

L'ordine dei coefficienti è importante in quanto corrisponde all'ordine crescente dei valori delle categorie della variabile fattore. Il primo coefficiente dell'elenco corrisponde al valore di gruppo minimo della variabile fattore e l'ultimo coefficiente corrisponde al valore massimo. Ad esempio, se sono presenti sei categorie della variabile fattore, i coefficienti –1, 0, 0, 0, 0,5 e 0,5 contrastano il primo gruppo con il quinto e il sesto gruppo. Per la maggior parte delle applicazioni la somma dei coefficienti deve essere 0. È possibile utilizzare anche insiemi la cui somma è diversa da 0, ma verrà visualizzato un messaggio di avvertimento.

ANOVA univariata: Test Post Hoc

Figura 23-3
Finestra di dialogo ANOVA univariata: Confronti multipli Post Hoc



Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervalli e i confronti a coppie multipli consentono di determinare quale media differisce dalle altre. I test a intervalli multipli consentono di identificare sottoinsiemi omogenei di medie che non differiscono le une dalle altre. Grazie ai confronti a coppie multipli è possibile verificare la differenza tra ciascuna coppia di medie e ottenere una matrice in cui gli asterischi indicano le medie di gruppo con differenze significative e un livello alfa 0,05.

Assumi varianze uguali

Il test delle differenze significative di Tukey, il GT2 di Hochberg, il test di Gabriel e il test di Scheffé sono test a confronti e intervalli multipli. Sono disponibili altri test di intervalli, ovvero *b* di Tukey, S-N-K (Student-Newman-Keuls), Duncan, R-E-G-W *F* (test *F* di

Ryan-Einot-Gabriel-Welsch), R-E-G-W *Q* (test a intervalli di Ryan-Einot-Gabriel-Welsch) e Waller-Duncan. I test a confronti multipli disponibili sono i seguenti: Bonferroni, test delle differenze significative di Tukey, Sidak, Gabriel, Hochberg, Dunnett, Scheffé e LSD (differenza meno significativa).

- **LSD.** Usa i test t per eseguire tutti i confronti a coppie tra medie di gruppo. Non viene apportata alcuna correzione al tasso di errore per i confronti multipli.
- **Bonferroni.** Usa i test t per eseguire confronti a coppie tra medie di gruppo, ma controlla il tasso di errore globale impostando il tasso di errore di ogni test al tasso di errore sperimentale diviso per il numero totale dei test. In questo modo il livello di significatività osservato è corretto tenendo conto che si stanno effettuando confronti multipli.
- **Sidak.** Test per confronti a coppie multipli basato sul test t. Effettua la correzione del livello di significatività per confronti multipli e fornisce una banda più stretta rispetto al test di Bonferroni.
- **Scheffé.** Effettua confronti congiunti a coppie simultanei per tutte le possibili coppie di medie. Utilizza la distribuzione di campionamento F. Può essere usato per esaminare tutte le possibili combinazioni lineari di medie di gruppo, non solo i confronti a coppie.
- **R-E-G-W F.** La procedura di Ryan-Einot-Gabriel-Welsch, basata su un test F.
- **R-E-G-W Q.** La procedura di Ryan-Einot-Gabriel-Welsch, basata su un intervallo studentizzato.
- **S-N-K.** Effettua tutti i confronti a coppie fra medie usando la distribuzione di intervallo studentizzata. Per dimensioni campionarie uguali, confronta anche coppie di medie entro sottoinsiemi omogenei, utilizzando una procedura pairwise. Le medie vengono ordinate dalla più alta alla più bassa e vengono verificate per prime le differenze estreme.
- **Tukey.** Usa la statistica di intervallo studentizzato per effettuare tutti i confronti a coppie tra gruppi. Imposta il tasso di errore sperimentale al valore del tasso di errore per l'insieme di tutti i confronti per coppie.
- **Tukey-b.** Utilizza la distribuzione di intervallo studentizzato per effettuare confronti a coppie tra gruppi. Il valore critico è la media fra il corrispondente valore per il test HSD di Tukey e quello del test di Student-Newman-Keuls.
- **Duncan.** Effettua confronti a coppie usando lo stesso ordine di confronti per passi usato nel test di Student Newman Keuls, impostando un livello di soglia per il tasso di errore valido per l'insieme dei test, invece di un tasso di errore diverso per ciascun test. Usa la statistica dell'intervallo studentizzato.
- **Hochberg (GT2).** Test per confronti multipli o per intervallo basato sul modulo studentizzato. Simile al test HSD di Tukey.
- **Gabriel.** Test per confronti a coppie basato sul modulo studentizzato, generalmente più indicato del test GT2 quando le celle hanno dimensioni diverse. Se la variabilità delle dimensioni delle celle risulta molto alta, il test di Gabriel può diventare poco conservativo.
- **Waller-Duncan.** Test per confronti multipli basato su una statistica t. Utilizza un approccio bayesiano.
- **Dunnett.** Test per confronti a coppie multipli basato sul test t, che confronta un insieme di trattamenti con un'unica media di controllo. L'ultima categoria è la categoria di controllo predefinita. In alternativa, è possibile scegliere la prima categoria. I test a due sensi consentono di verificare che la media in qualsiasi livello del fattore (ad eccezione della categoria di controllo) non sia uguale a quella della categoria di controllo. I test di <controllo

consentono di verificare se la media di qualsiasi livello del fattore sia minore di quella della categoria di controllo. I test di controllo > consentono di verificare se la media di qualsiasi livello del fattore sia maggiore di quella della categoria di controllo.

Non assumere varianze uguali

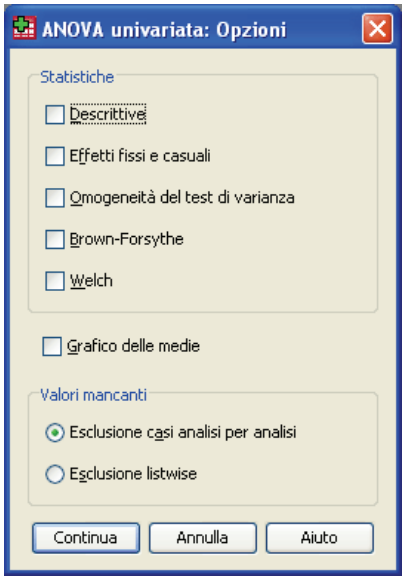
Sono disponibili test a confronti multipli che non ipotizzano varianze uguali, ovvero T2 di Tamhane, T3 di Dunnett, Games-Howell e *C* di Dunnett.

- **Tamhane (T2).** Test per confronti a coppie basato sul test t, appropriato quando le varianze non sono uguali.
- **Dunnett (T3).** Test per confronti a coppie basato sul modulo studentizzato. appropriato quando le varianze non sono uguali.
- **Games-Howell.** Test per confronti a coppie non molto conservativo, appropriato quando le varianze non sono uguali.
- **C di Dunnett.** Test per confronti a coppie basato sull'intervallo studentizzato, appropriato quando le varianze non sono uguali.

Nota: Per interpretare più rapidamente l'output dei test post hoc è consigliabile deselezionare Nascondi righe e colonne vuote nella finestra di dialogo Proprietà tabella (in una tabella pivot attivata, scegliere Proprietà tabella dal menu Formato).

ANOVA univariata: Opzioni

Figura 23-4
Finestra di dialogo ANOVA univariata: Opzioni



Statistiche. Consente di scegliere una o più delle seguenti opzioni:

- **Descrittive.** Consente di calcolare il numero di casi, la media, la deviazione standard, l'errore standard della media, il valore minimo e massimo e gli intervalli di confidenza al 95% per ciascuna variabile dipendente di ciascun gruppo.

- **Effetti fissi e casuali.** Consente di visualizzare la deviazione standard, l'errore standard e l'intervallo di confidenza del 95% per il modello degli effetti fissi e l'errore standard, l'intervallo di confidenza del 95% e la stima della varianza tra componenti per il modello degli effetti casuali.
- **Omogeneità del test di varianza.** Consente di calcolare il test di Levene per verificare l'uguaglianza tra gruppi di variabili. Questo test non si basa sull'ipotesi di normalità.
- **Brown-Forsythe.** Consente di calcolare la statistica di Brown-Forsythe per verificare l'uguaglianza delle medie dei gruppi. Questa statistica è preferibile alla statistica *F* nel caso in cui non sia valida l'ipotesi di uguaglianza della varianza.
- **Welch.** Consente di calcolare la statistica di Welch per verificare l'uguaglianza delle medie dei gruppi. Questa statistica è preferibile alla statistica *F* nel caso in cui non sia valida l'ipotesi di uguaglianza della varianza.

Grafico delle medie. Consente di visualizzare un grafico che rappresenta le medie dei sottogruppi (le medie di ciascun gruppo definite dai valori della variabile fattore).

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi analisi per analisi.** I casi con valori mancanti per la variabile dipendente o fattore per una particolare analisi non verranno utilizzati in tale analisi. Non verranno utilizzati nemmeno i casi che non rientrano nell'intervallo specificato per la variabile fattore.
- **Esclusione listwise.** I casi con valori mancanti per la variabile fattore o qualsiasi variabile dipendente inclusa nella lista delle variabili dipendenti nella finestra di dialogo principale verranno esclusi da tutte le analisi. Se non sono state specificate più variabili dipendenti, l'opzione non produrrà alcun effetto.

Opzioni aggiuntive del comando ONEWAY

Il linguaggio della sintassi dei comandi consente inoltre di:

- Ottenere statistiche con effetti fissi e casuali. Deviazione standard, errore standard della media e intervalli di confidenza al 95% per il modello con effetti fissi. Errore standard, intervalli di confidenza al 95% e stima della varianza dei componenti per il modello con effetti casuali (con il sottocomando `STATISTICS=EFFECTS`).
- Specificare i livelli alfa per la differenza meno significativa nei test per confronti a coppie multipli di Bonferroni, Duncan e Scheffé (con il sottocomando `RANGES`).
- Scrivere una matrice di medie, deviazioni standard e frequenze, oppure leggere una matrice di medie, frequenze, varianze raggruppate e gradi di libertà per le varianze raggruppate. Queste matrici possono essere usate al posto dei dati grezzi per effettuare un'analisi univariata della varianza (con il sottocomando `MATRIX`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
24

Analisi GLM univariato

La procedura GLM univariato consente di eseguire un'analisi di regressione e un'analisi della varianza per una variabile dipendente tramite uno o più fattori e/o variabili. Le variabili fattore suddividono la popolazione in gruppi. Con questa procedura GLM (modello lineare generalizzato, General Linear Model) è possibile verificare ipotesi nulle relative agli effetti di altre variabili sulle medie di vari raggruppamenti di una sola variabile dipendente. È possibile analizzare le interazioni tra fattori e gli effetti di singoli fattori, alcuni dei quali possono essere casuali. È inoltre possibile includere gli effetti delle covariate e le interazioni tra covariate e fattori. Nell'analisi di regressione, le variabili indipendenti (stimatori) vengono specificate come covariate.

È possibile verificare sia modelli bilanciati che modelli non bilanciati. Un disegno è bilanciato se ciascuna cella del modello include lo stesso numero di casi. Oltre alla verifica delle ipotesi, la procedura GLM univariato consente di ottenere stime dei parametri.

Per la verifica di ipotesi sono disponibili contrasti a priori usati di frequente. Dopo che da un test *F* globale è risultata una certa significatività, è inoltre possibile eseguire test post hoc per valutare le differenze tra medie specifiche. L'opzione Medie marginali stimate consente di ottenere stime dei valori medi previsti delle celle incluse nel modello. I grafici di profilo, o grafici di interazione, di tali medie consentono di visualizzare in modo semplice alcune delle relazioni.

Residui, valori attesi, distanza di Cook e valori d'influenza possono essere salvati come variabili nel file di dati per la verifica di ipotesi.

Minimi quadrati ponderati consente di specificare una variabile per l'assegnazione di pesi diversi alle osservazioni per un'analisi di minimi quadrati ponderati (WLS), in alcuni casi per compensare la diversa precisione della misura.

Esempio. Per diversi anni vengono raccolti i dati relativi ai singoli partecipanti alla maratona di Chicago. Il tempo impiegato da ciascun partecipante per completare la maratona è la variabile dipendente. Altri fattori presi in considerazione sono le condizioni metereologiche (freddo, caldo o temperatura moderata), il numero di mesi di allenamento, il numero di maratone corse in precedenza e il sesso. L'età è considerata una covariata. Dallo studio può risultare che il sesso rappresenta un effetto significativo, così come l'interazione tra sesso e condizioni meteorologiche.

Metodi. Per la valutazione di ipotesi diverse è possibile usare la somma dei quadrati Tipo I, Tipo II, Tipo III e Tipo IV. Il metodo predefinito è il Tipo III.

Statistiche. I seguenti test post hoc di intervalli e confronti multipli: Differenza meno significativa (LSD), Bonferroni, Sidak, Scheffé, Ryan-Einot-Gabriel-Welsch multiplo basato su test *F*, Ryan-Einot-Gabriel-Welsch a intervallo multiplo, Student-Newman-Keuls, differenze significative di Tukey, *b* di Tukey, Duncan, Hochberg (GT2), Gabriel, *t* di Waller-Duncan, Dunnett (a una e a due vie), Tamhane (T2), Dunnett (T3), Games-Howell e *C* di Dunnett. Statistiche descrittive: medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle. Test di Levene di omogeneità della varianza.

Grafici. Grafici di variabilità vs. densità, dei residui e di profilo (interazione).

Dati. La variabile dipendente è quantitativa. I fattori sono categoriali. Vi possono essere associati valori numerici o valori stringa composti da un massimo di otto caratteri. Le covariate sono variabili quantitative correlate alla variabile dipendente.

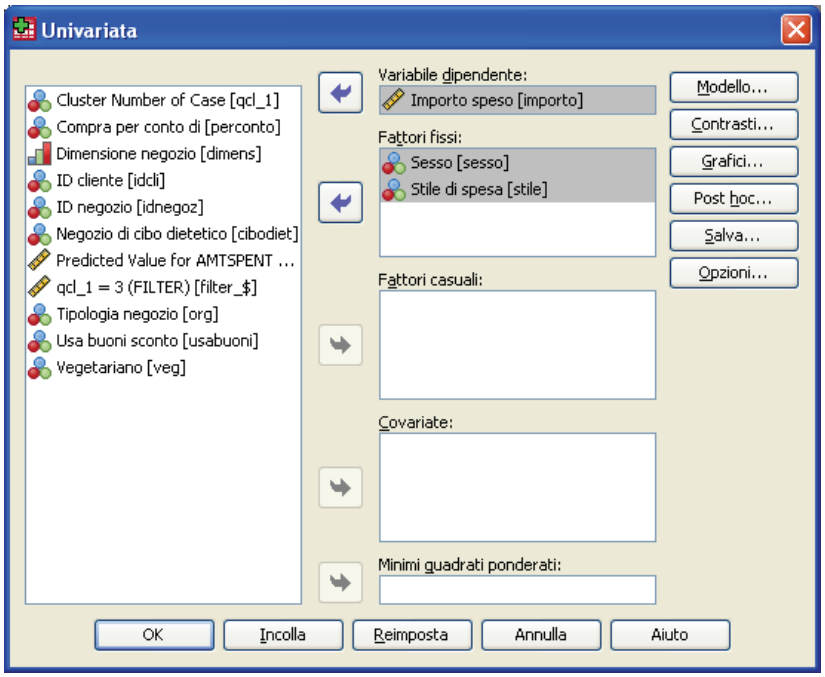
Assunzioni. I dati sono costituiti da un campione casuale derivato da una popolazione normale in cui tutte le varianze di cella sono uguali. L'analisi della varianza è uno stimatore robusto degli scostamenti dalla normalità, anche se i dati devono essere simmetrici. Per la verifica di ipotesi, è possibile usare test di omogeneità della varianza e i grafici di variabilità vs. intensità. È inoltre possibile esaminare residui e grafici dei residui.

Per ottenere tabelle di GLM univariato

▶ Dai menu, scegliere:

Analizza
  Modello lineare generalizzato
    Univariata...

Figura 24-1
Finestra di dialogo GLM univariato



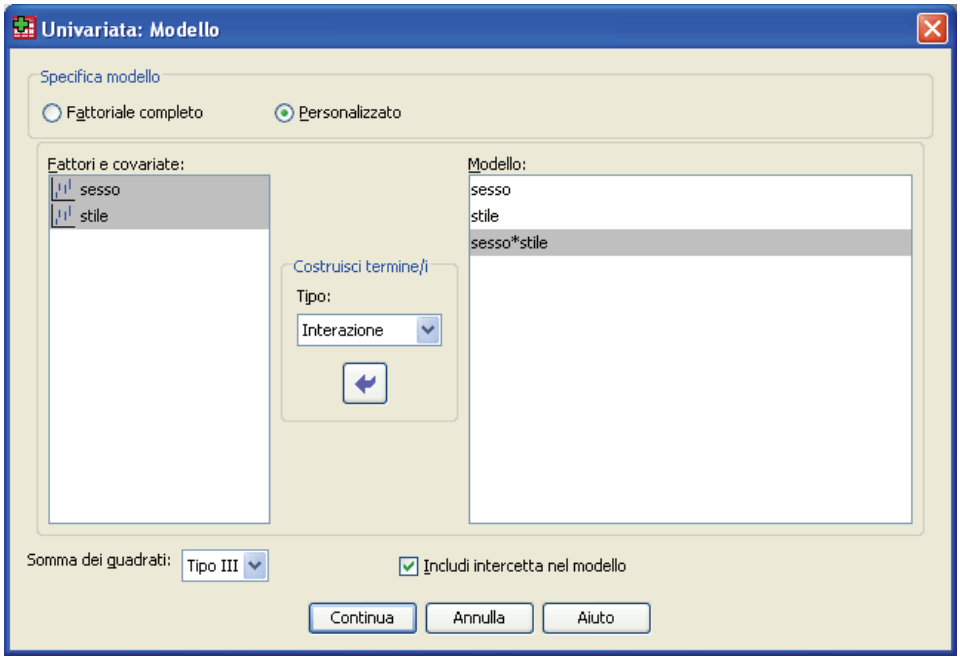
▶ Selezionare una variabile dipendente.

▶ Selezionare le variabili per l'opzione Fattori fissi, Fattori casuali o Covariate, a seconda dei dati in uso.

▶ È inoltre possibile utilizzare l'opzione Minimi quadrati ponderati per specificare l'analisi dei minimi quadrati ponderati. Casi con valore 0, negativo o mancante per la variabile di ponderazione saranno esclusi dall'analisi. Non è possibile utilizzare come variabile di ponderazione una variabile già inclusa nel modello.

GLM - Univariato: Modello

Figura 24-2
Finestra di dialogo Univariata: Modello



Specifica modello. Un modello fattoriale completo contiene tutti gli effetti principali dei fattori e delle covariate e tutte le interazioni fattore per fattore. Non contiene interazioni di covariate. Selezionare Personalizzato per specificare un solo sottoinsieme di interazioni o interazioni tra fattori e covariate. È necessario indicare tutti i termini da includere nel modello.

Fattori e covariate. I fattori e le covariate sono elencati.

Modello. Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato Personalizzato, è possibile selezionare gli effetti principali e le interazioni desiderate per l'analisi da eseguire.

Somma dei quadrati. Metodo per il calcolo della somma dei quadrati. Il metodo Somma dei quadrati in genere utilizzato con modelli bilanciati o non bilanciati privi di celle mancanti è il Tipo III.

Includi l'intercetta nel modello. L'intercetta viene in genere inclusa nel modello. Se è possibile presumere che i dati passino attraverso l'origine, l'intercetta può essere esclusa.

Costruisci termini

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti 2-vie. Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti 3-vie. Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti 4-vie. Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti 5-vie. Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Somma dei quadrati

Per il modello è possibile scegliere un tipo di somma dei quadrati. Il Tipo III, il tipo di default, è quello usato più di frequente.

Tipo I. Questo metodo è definito anche scomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene corretto solo per i termini del modello che lo precedono. Il metodo Somma dei quadrati Tipo I è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato in cui gli effetti principali vengono specificati prima degli effetti di interazione di ordine 1, ciascuno dei quali viene a sua volta specificato prima degli effetti di interazione di ordine 2 e così via.
- Un modello di regressione polinomiale in cui qualsiasi termine di ordine più basso è specificato prima dei termini di ordine più elevato.
- Un modello nidificato in modo puro in cui il primo effetto specificato è nidificato nel secondo, il quale è a sua volta nidificato nel terzo e così via. Questo tipo di nidificazione può essere specificato esclusivamente tramite la sintassi.

Tipo II. Questo metodo consente di calcolare le somme dei quadrati di un effetto del modello corretto per tutti gli altri effetti "appropriati". È considerato appropriato un effetto corrispondente a tutti gli effetti che non includono l'effetto in esame. Il metodo Somma dei quadrati Tipo II è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato.
- Qualsiasi modello che include solo effetti principali del fattore.
- Qualsiasi modello di regressione.
- Un disegno nidificato in modo puro. Questo tipo di nidificazione può essere specificato tramite la sintassi.

Tipo III. Tipo di default. Questo metodo consente di calcolare le somma dei quadrati di un effetto del disegno come la somma dei quadrati corretta per qualsiasi altro effetto che non lo include e ortogonale rispetto agli eventuali effetti che lo contengono. Il vantaggio associato a questo tipo di somme dei quadrati è che non varia al variare delle frequenze di cella, a condizione che la forma generale di stimabilità rimanga costante. È pertanto considerato utile per modelli non bilanciati privi di celle mancanti. In un disegno fattoriale privo di celle mancanti, questo metodo equivale alla tecnica dei quadrati delle medie ponderate di Yates. Il metodo Somma dei quadrati Tipo III è in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato o non bilanciato e privo di celle vuote.

Tipo IV. Questo metodo è specifico per situazioni con celle mancanti. Per qualsiasi effetto *F* del disegno, se *F* non è incluso in nessun altro effetto, allora Tipo IV = Tipo III = Tipo II. Se invece *F* è incluso in altri effetti, con il Tipo IV i contrasti creati tra i parametri in *F* vengono distribuiti

equamente tra tutti gli effetti di livello superiore. Il metodo Somma dei quadrati Tipo IV viene in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato e non bilanciato contenente celle vuote.

GLM - Univariato: Contrasti

Figura 24-3
Finestra di dialogo Univariata: Contrasti

I contrasti consentono di verificare il grado di differenza tra i livelli di un fattore. È possibile specificare un contrasto per ciascun fattore del modello (in un modello a misure ripetute, un contrasto per ciascun fattore tra soggetti). I contrasti rappresentano combinazioni lineari dei parametri.

La verifica di ipotesi è basata sull'ipotesi nulla **LB** = 0, dove **L** è la matrice dei coefficienti di contrasto e **B** è il vettore dei parametri. Quando si specifica un contrasto, viene creata una matrice **L**. Le colonne della matrice **L** corrispondenti al fattore corrispondono al contrasto. Le altre colonne vengono corrette in modo che la matrice **L** possa essere stimata.

L'output include una statistica *F* per ciascun insieme di contrasti. Per le differenze dei contrasti vengono inoltre visualizzati gli intervalli di confidenza simultanei di tipo Bonferroni basati su una distribuzione *t* di Student.

Contrasti disponibili

Sono disponibili i contrasti deviazione, semplici, differenza, Helmert, ripetuti e polinomiali. Per i contrasti deviazione e i contrasti semplici, è possibile stabilire se la categoria di riferimento corrisponde alla prima o all'ultima categoria.

Tipi di contrasto

Deviazione. Consente di confrontare la media di ciascun livello, a eccezione di una categoria di riferimento, con la media di tutti i livelli (media principale). L'ordine dei livelli del fattore può essere un ordine qualsiasi.

Semplice. Consente di confrontare la media di ciascun livello con la media di un livello specifico. Questo tipo di contrasto risulta utile quando è disponibile un gruppo di controllo. Come categoria di riferimento, è possibile scegliere la prima o l'ultima categoria.

Differenza. Consente di confrontare la media di ciascun livello (a eccezione del primo) con la media dei livelli precedenti. Questo tipo di contrasto è a volte definito contrasto inverso di Helmert.

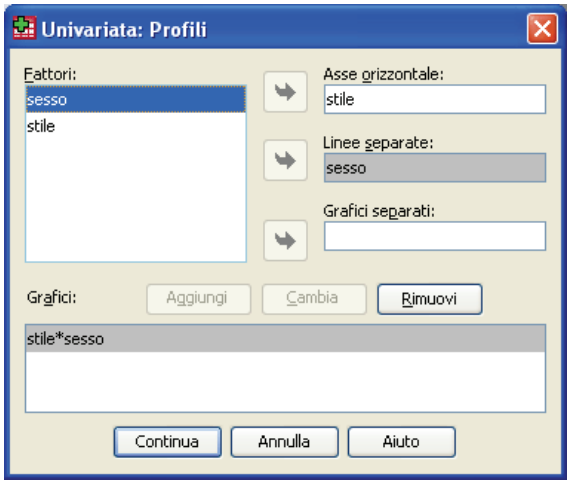
Helmert. Consente di confrontare la media di ciascun livello del fattore (a eccezione dell'ultimo) con la media dei livelli successivi.

Ripetuto. Consente di confrontare la media di ciascun livello (a eccezione dell'ultimo) con la media del livello successivo.

Polinomiale. Consente di confrontare l'effetto lineare, quadratico, cubico e così via. Tutte le categorie del primo grado di libertà includono l'effetto lineare, quelle del secondo includono l'effetto quadratico e così via. Questi contrasti sono spesso usati per la stima di trend polinomiali.

GLM - Univariato: Profili

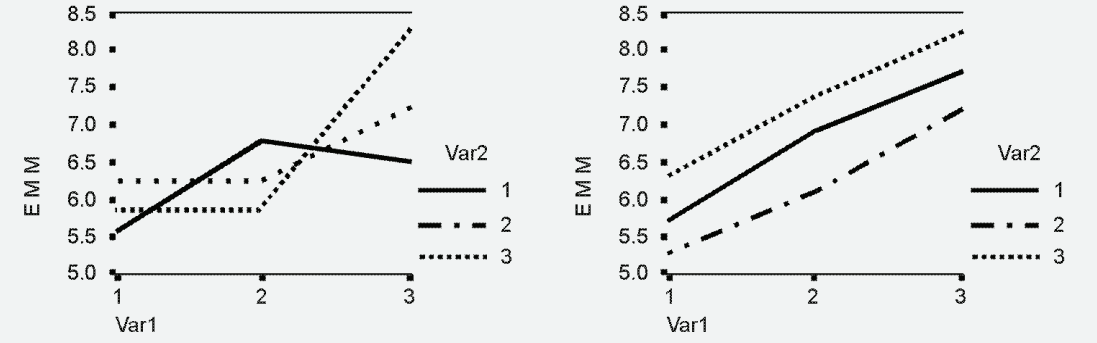
Figura 24-4
Finestra di dialogo Univariata: Profili



I profili, o grafici di interazione, risultano utili per il confronto delle medie marginali di un modello. Un profilo è un grafico lineare in cui ciascun punto indica la media marginale stimata di una variabile dipendente (corretta per le covariate) in corrispondenza di un solo livello di un fattore. È possibile utilizzare i livelli di un secondo fattore per creare linee distinte. È possibile utilizzare ciascun livello di un terzo fattore per creare un grafico distinto. Tutti gli eventuali fattori casuali e fissi sono disponibili per i grafici. In analisi multivariate i grafici di profilo vengono creati per ciascuna variabile dipendente. Nei grafici di profilo per un'analisi a misure ripetute è possibile includere sia fattori tra soggetti che fattori entro soggetti. GLM Multivariato e GLM Misure ripetute sono disponibili solo se è stata installata l'opzione Advanced Statistics.

Il profilo di un fattore mostra se le medie marginali stimate aumentano o diminuiscono tra i vari livelli. Nel caso di due o più fattori, le linee parallele indicano che tra i fattori non esiste alcuna interazione, ovvero che è possibile analizzare i livelli di un solo fattore. Le linee che si incrociano indicano invece che esiste un'interazione.

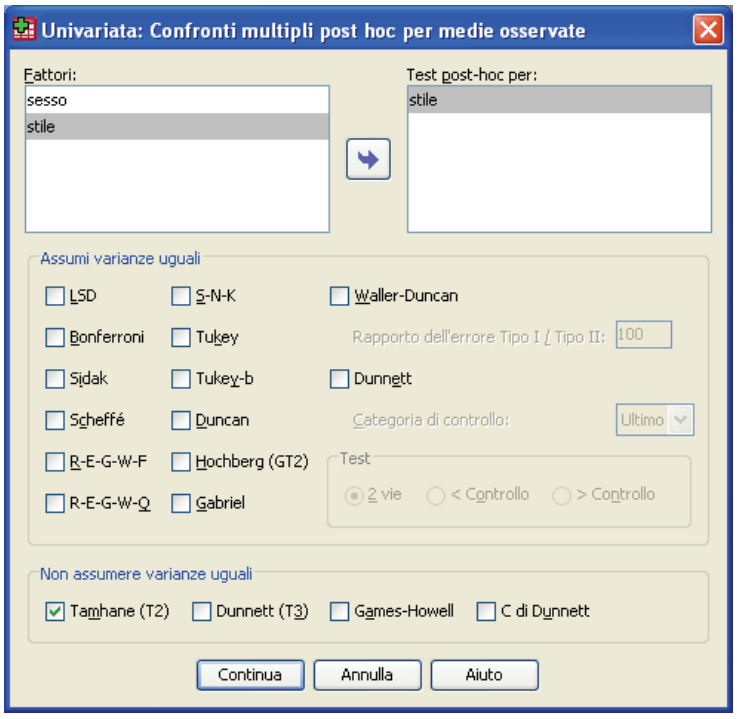
Figura 24-5
Grafico con linee non parallele (a sinistra) e grafico con linee parallele (a destra)



I grafici definiti tramite la selezione dei fattori per l'asse orizzontale e, se lo si desidera, dei fattori di linee e di grafici separati devono essere inclusi nell'elenco dei grafici.

GLM - Univariato: Confronti post hoc

Figura 24-6
Finestra di dialogo Post Hoc



Test per confronti multipli post hoc. Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervalli e i confronti a coppie multipli consentono di determinare quale media differisce dalle altre. I confronti vengono eseguiti su valori a cui non è stata apportata alcuna correzione. Questi test vengono usati solo con fattori fissi tra soggetti. Nella procedura GLM a misure ripetute, questi test non sono disponibili se non sono presenti fattori tra soggetti e i test per confronti multipli post hoc vengono eseguiti per la media tra i livelli dei fattori entro soggetti.

Per la procedura GLM multivariato, i test post hoc vengono eseguiti separatamente per ciascuna variabile dipendente. GLM Multivariato e GLM Misure ripetute sono disponibili solo se è stata installata l'opzione Advanced Statistics.

I test per confronti multipli usati più di frequente sono il test di Bonferroni e il test delle differenze significative di Tukey. Il **test di Bonferroni**, basato sulla statistica *t* di Student, consente di correggere il livello di significatività osservato in base al fatto che vengono eseguiti confronti multipli. Il **test t di Sidak** corregge inoltre il livello di significatività ed è più restrittivo del test di Bonferroni. Il **test delle differenze significative di Tukey** utilizza la statistica di intervallo studentizzato per effettuare tutti i confronti a coppie tra gruppi e imposta il tasso di errore sperimentale sul valore del tasso di errore per l'insieme di tutti i confronti per coppie. Quando si eseguono test su un elevato numero di coppie di medie, il test delle differenze significative di Tukey risulta più efficace rispetto al test di Bonferroni. Nel caso di un numero limitato di coppie, risulta invece più efficace il test di Bonferroni.

Il test di **Hochberg (GT2)** è simile al test delle differenze significative di Tukey, ma utilizza il modulo massimo studentizzato. Il test di Tukey risulta in genere più efficace. Anche il **test dei confronti a coppie di Gabriel** utilizza il modulo massimo studentizzato ed è in genere più indicativo del test di Hochberg (GT2) quando le dimensioni delle celle sono diverse. Se la variabilità delle dimensioni delle celle risulta molto alta, il test di Gabriel può diventare poco conservativo.

Il **test t per confronti multipli a coppie di Dunnett** confronta un insieme di trattamenti con una media di controllo singola. L'ultima categoria è la categoria di controllo predefinita. In alternativa, è possibile scegliere la prima categoria. È inoltre possibile scegliere un test a 2 vie oppure a 1 via. Per verificare che la media in qualsiasi livello del fattore (a eccezione della categoria di controllo) non sia uguale a quella della categoria di controllo, è necessario usare un test a due sensi. Per verificare se la media di qualsiasi livello del fattore è minore di quella della categoria di controllo, selezionare < Controllo. In modo analogo, per verificare se la media di qualsiasi livello del fattore è maggiore di quella della categoria di controllo, selezionare > Control.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) hanno sviluppato due test a intervalli decrescenti multipli. Le procedure a multipli decrescenti verificano in primo luogo se tutte le medie sono uguali. Se le medie non risultano tutte uguali, il test di uguaglianza viene eseguito su un sottoinsieme di medie. Il test **R-E-G-W F** è basato su un test *F*, mentre **R-E-G-W Q** è basato sull'intervallo studentizzato. Questi test risultano più efficaci rispetto ai test a intervallo multiplo di Duncan e Student-Newman-Keuls, che sono pure procedure a intervalli decrescenti multipli. È tuttavia consigliabile non usarli con celle di dimensioni non uguali.

Quando le varianze non sono uguali, è necessario utilizzare il test **Tamhane (T2)** (test per confronti a coppie conservativo basato su un test *t*), il test di **Dunnett T3** (test per confronti a coppie basato sul modulo studentizzato), il test per confronti a coppie di **Games-Howell** (a volte poco conservativo) o il test **C di Dunnett** (test per confronti a coppie basato sull'intervallo studentizzato). Notare che questi test non sono validi e non verranno eseguiti se il modello contiene più fattori.

Il test a intervallo multiplo di Duncan, il test di Student-Newman-Keuls (**S-N-K**) e il test **b di Tukey** sono test per intervallo che classificano le medie raggruppate e calcolano un valore di intervallo. Questi test sono usati meno frequentemente dei test descritti in precedenza.

Il **test t di Waller-Duncan** utilizza un approccio bayesiano. Si tratta di un test a intervallo che usa la media armonica della dimensione campione nel caso di dimensioni campione non uguali.

Il livello di significatività del test di **Scheffé** consente la verifica di tutte le possibili combinazioni lineari delle medie di gruppo e non dei soli confronti a coppie disponibili in questa funzione. Il test di Scheffé risulta pertanto più conservativo rispetto ad altri test, ovvero per ottenere un livello sufficiente di significatività, è richiesta una differenza maggiore tra le medie.

Il test per confronti a coppie multipli Differenza meno significativa o **LSD**, è equivalente a più test *t* tra tutte le coppie di gruppi. Lo svantaggio di questo test è che non viene eseguito alcun tentativo di correzione del livello di significatività osservata per confronti multipli.

Test visualizzati. I confronti a coppie sono disponibili per i test LSD, Sidak, Bonferroni, Games-Howell, Tamhane (T2) e (T3), test *C* di Dunnett e Dunnett (T3). Per i test S-N-K, *b* di Tukey, Duncan, R-E-G-W *F*, R-E-G-W *Q* e Waller sono disponibili sottoinsiemi omogenei per test per intervallo. Il test delle differenze significative di Tukey, i test Hochberg (GT2), Gabriel e Scheffé sono sia test per confronti multipli che test a intervallo.

GLM - Univariato: Salva

Figura 24-7
Finestra di dialogo Salva

È possibile salvare i valori attesi dal modello, le misure correlate e i residui come nuove variabili nell'Editor dei dati. Molte di queste variabili possono essere usate per l'esame di ipotesi sui dati. Per salvare i valori in modo da poterli usare in un'altra sessione SPSS Statistics, è necessario salvare il file di dati corrente.

Valori attesi. Valori attesi dal modello per ciascun caso.

- **Non standardizzati.** I valori risultanti dal modello per la variabile dipendente e per ciascun caso.
- **Pesati.** I valori attesi ponderati non standardizzati. Disponibile solo se è stata selezionata una variabile WLS.
- **Errore standard.** Una stima della deviazione standard del valore medio della variabile dipendente per i casi che hanno gli stessi valori delle variabili indipendenti.

Diagnostici. Misure per l'identificazione dei casi con combinazioni di valori insolite per le variabili indipendenti e dei casi che possono avere una notevole influenza sul modello.

- **Distanza di Cook.** Una misura di quanto cambierebbero i residui di tutti i casi se un particolare caso fosse escluso dal calcolo dei coefficienti di regressione. Valori alti indicano che l'esclusione di un caso dal calcolo dei coefficienti di regressione ne modificherebbe sostanzialmente il valore.
- **Valori di influenza.** Valori di influenza non centrati. Una misura dell'influenza di ciascun caso sull'adattamento di un modello di regressione.

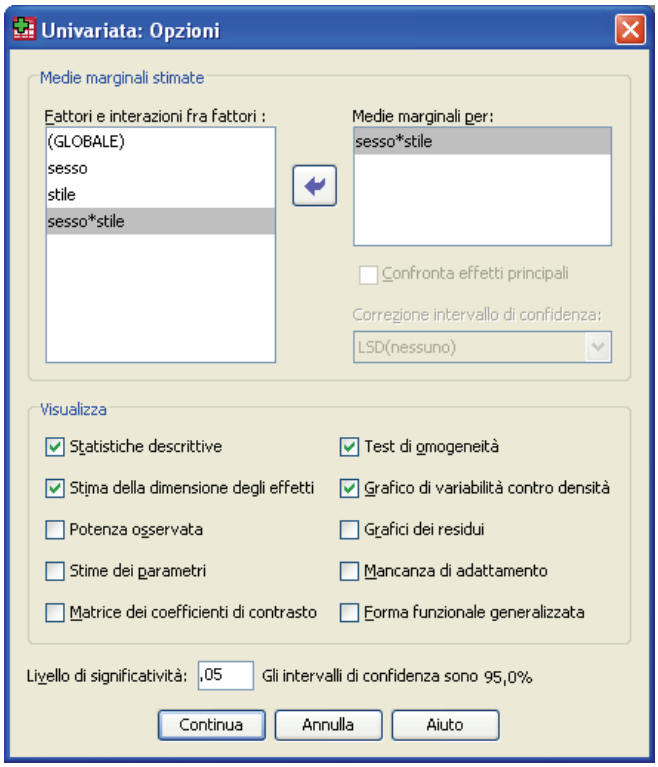
Residui. Un residuo non standardizzato corrisponde al valore effettivo della variabile dipendente diminuito del valore atteso dal modello. Sono inoltre disponibili residui standardizzati, studentizzati ed eliminati. Se è stata selezionata una variabile WLS, saranno inoltre disponibili residui non standardizzati pesati.

- **Non standardizzati.** La differenza tra un valore osservato e il valore stimato dal modello.
- **Pesati.** I residui ponderati non standardizzati. Disponibile solo se è stata selezionata una variabile WLS.
- **Standardizzati.** Il residuo diviso per una stima della deviazione standard. Il residuo standardizzato, conosciuto anche come residuo di Pearson, ha media 0 e deviazione standard 1.
- **Studentizzati.** Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, a seconda della distanza tra i valori assunti per questo caso dalle variabili indipendenti e le medie delle variabili indipendenti.
- **Per cancellazione.** Il residuo per un caso se quel caso venisse escluso dal calcolo dei coefficienti di regressione. È la differenza tra il valore della variabile dipendente e il valore stimato corretto.

Statistiche dei coefficienti. Scrive una matrice della varianza-covarianza delle stime dei parametri nel modello in un nuovo file di dati della sessione attiva o in un file di dati SPSS Statistics esterno. Per ciascuna variabile dipendente è inoltre disponibile una riga di stime dei parametri, una riga di valori di significatività per le statistiche *t* corrispondenti alle stime dei parametri e una riga di gradi di libertà dei residui. Per ciascuna variabile dipendente di modelli multivariati sono disponibili righe simili. Il file matrice può essere usato in altre procedure che leggono i file matrice.

GLM - Univariato: Opzioni

Figura 24-8
Finestra di dialogo Opzioni



In questa finestra di dialogo sono disponibili statistiche opzionali. Le statistiche vengono calcolate tramite un modello di effetti fissi.

Medie marginali stimate. Selezionare i fattori e le interazioni per cui si desiderano le stime delle medie marginali della popolazione nelle celle. Queste medie vengono corrette per le eventuali covariate.

- **Confronta effetti principali.** Consente di eseguire confronti a coppie senza correzione tra le medie marginali stimate di qualsiasi effetto principale del modello, per fattori sia tra soggetti che entro soggetti. Questa opzione è disponibile solo se nell'elenco Medie marginali per sono stati selezionati effetti principali.
- **Correzione intervallo di confidenza.** Selezionare la differenza meno significativa (LSD), la correzione di Bonferroni o di Sidak agli intervalli di confidenza e la significatività. Questo comando è disponibile solo se è stato selezionato Confronta effetti principali.

Visualizzazione. Selezionare Statistiche descrittive per produrre medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle. La funzione Stima della dimensione degli effetti fornisce un valore eta-quadrato parziale per ciascun effetto e per ciascuna stima dei parametri. La statistica eta-quadrato consente di ottenere la proporzione della variabilità totale attribuibile a un fattore. Selezionare Potenza osservata per ottenere la potenza del test nel caso in cui l'ipotesi alternativa sia basata sul valore osservato. Selezionare Stime dei parametri per

ottenere stime dei parametri, errori standard, test *t*, intervalli di confidenza e la potenza osservata per ciascun test. Selezionare Matrice dei coefficienti di contrasto per ottenere la matrice **L**.

La funzione Test di omogeneità produce il test di Levene per l'omogeneità della varianza per ogni variabile dipendente su tutte le combinazioni di livello dei fattori fra soggetti, solo per i fattori fra soggetti. Le opzioni Grafici di variabilità vs. densità e Grafici dei residui risultano utili per la verifica di ipotesi sui dati. Se non è disponibile alcun fattore, questa opzione risulta disattivata. Selezionare Grafici dei residui per ottenere un grafico dei residui osservati, attesi e standardizzati per ciascuna variabile dipendente. Questi grafici risultano utili per l'analisi dell'ipotesi di uguaglianza della varianza. Selezionare Mancanza di adattamento per controllare che la relazione fra la variabile dipendente e le variabili indipendenti possa essere descritta in modo adeguato dal modello. La funzione Forma funzionale generalizzata consente di creare test di ipotesi personalizzati basati sulla forma funzionale generalizzata. Le righe di una matrice dei coefficienti di contrasto sono combinazioni lineari della forma funzionale generalizzata.

Livello di significatività. Potrebbe risultare utile correggere il livello di significatività usato nei test post hoc e il livello di confidenza usato per la costruzione degli intervalli di confidenza. Il valore specificato viene inoltre usato per il calcolo della potenza osservata per il test. Quando si specifica un livello di significatività, nella finestra di dialogo viene visualizzato il livello di intervalli di confidenza associato.

Funzioni aggiuntive del comando UNIANOVA

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare gli effetti nidificati del disegno (tramite il sottocomando `DESIGN`).
- Specificare test di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando `TEST`).
- Specificare contrasti multipli (tramite il sottocomando `CONTRAST`).
- Includere valori mancanti definiti dall'utente (tramite il sottocomando `MISSING`).
- Specificare criteri EPS (tramite il sottocomando `CRITERIA`).
- Costruire una matrice **L**, **M** o **K** personalizzata (tramite il sottocomando `LMATRIX`, `MMATRIX` o `KMATRIX`).
- Per i contrasti deviazione e i contrasti semplici, specificare una categoria di riferimento intermedia (tramite il sottocomando `CONTRAST`).
- Specificare metrica per contrasti polinomiali (tramite il sottocomando `CONTRAST`).
- Specificare termini di errore per confronti post-hoc (tramite il sottocomando `POSTHOC`).
- Calcolare medie marginali stimate per qualsiasi fattore o interazione tra fattori per i fattori elencati (tramite il sottocomando `EMMEANS`).
- Assegnare un nome alle variabili temporanee (tramite il sottocomando `SAVE`).
- Costruire un file di dati di matrici di correlazione (tramite il sottocomando `OUTFILE`).
- Costruire un file di dati di matrici contenente statistiche derivate dai dati della tabella ANOVA tra soggetti (tramite il sottocomando `OUTFILE`).
- Salvare la matrice del disegno in un nuovo file di dati (tramite il sottocomando `OUTFILE`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
25

Correlazioni bivariate

La procedura Correlazioni bivariate consente di calcolare coefficiente di correlazione di Pearson, rho di Spearman e tau-*b* di Kendall con i rispettivi livelli di significatività. Le correlazioni consentono di misurare la relazione tra variabili o punteggi di rango. Prima di calcolare un coefficiente di correlazione, è necessario valutare la presenza di valori anomali nei dati (che possono causare risultati errati) e l'esistenza di una relazione lineare. Il coefficiente di correlazione di Pearson è una misura di associazione lineare. Due variabili possono essere perfettamente correlate, ma se la relazione non è lineare il coefficiente di correlazione di Pearson non è la statistica migliore per misurare tale associazione.

Esempio. Il numero di partite vinte da una squadra di baseball è correlato con la media dei punti totalizzati per ciascuna partita? Un grafico a dispersione indica l'esistenza di una relazione lineare. Dall'analisi dei dati relativi alla stagione NBA 1994–1995 risulta che il coefficiente di correlazione di Pearson (0,581) è significativo al livello 0,01. Si può presumere che il numero di partite vinte per stagione sia inversamente proporzionale ai punti totalizzati dagli avversari. Queste variabili sono legate da una correlazione negativa (–0,401), significativa al livello 0,05.

Statistiche. Per ogni variabile: numero di casi con valori non mancanti, media e deviazione standard. Per ciascuna coppia di variabili: coefficiente di correlazione di Pearson, rho di Spearman, tau-*b* di Kendall, prodotto incrociato delle deviazioni, covarianza.

Dati. Utilizzare le variabili quantitative simmetriche per il coefficiente di correlazione di Pearson e le variabili quantitative o le variabili con categorie ordinate per rho di Spearman e tau-*b* di Kendall.

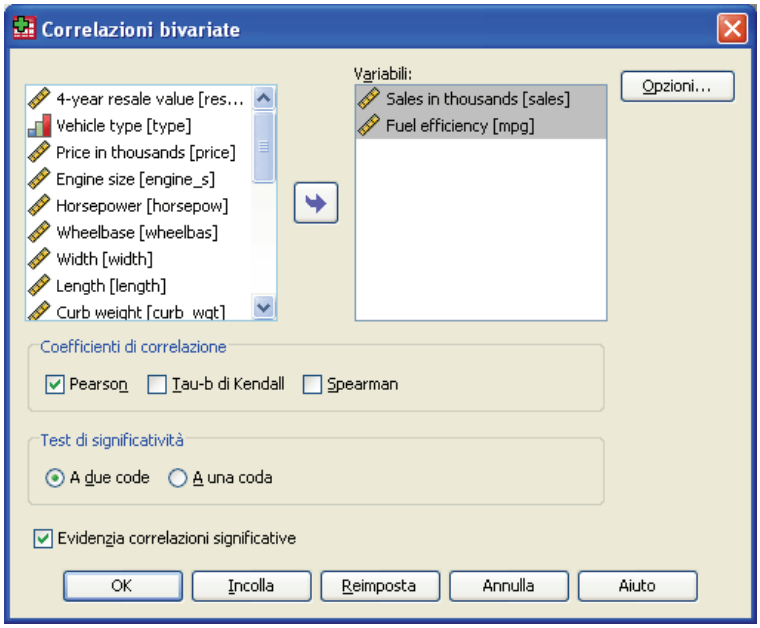
Assunzioni. Il coefficiente di correlazione di Pearson assume che ciascuna coppia di variabili sia bivariata normale.

Per ottenere correlazioni bivariate

Dai menu, scegliere:

Analizza
  Correlazione
    Bivariata...

Figura 25-1
Finestra di dialogo Correlazioni bivariate



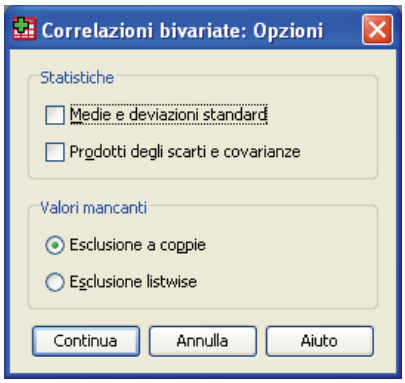
► Selezionare una o più variabili numeriche.

Sono inoltre disponibili le seguenti opzioni:

- **Coefficienti di correlazione.** Per variabili quantitative normalmente distribuite, scegliere il coefficiente di correlazione di Pearson. Se i dati non sono normalmente distribuiti o prevedono categorie ordinate, scegliere Tau-b di Kendall o Spearman, per misurare l'associazione tra punteggi di rango. I valori dei coefficienti di correlazione variano da –1 (relazione negativa perfetta) a +1 (relazione positiva perfetta). Il valore 0 indica l'assenza di relazione lineare. Interpretando i risultati, evitare di trarre conclusioni di tipo causa-effetto sulla base di una correlazione significativa.
- **Test di significatività.** È possibile selezionare le probabilità a una coda o a due code. Se si conosce in anticipo la direzione dell'associazione, selezionare A una coda.. In alternativa, selezionare A due code.
- **Evidenzia correlazioni significative.** I coefficienti di correlazione significativi al livello 0,05 vengono identificati con un asterisco singolo e quelli significativi al livello 0,01 con due asterischi.

Correlazioni bivariate: Opzioni

Figura 25-2
Finestra di dialogo Correlazioni bivariate: Opzioni



Statistiche. Per le correlazioni di Pearson, è possibile scegliere una o entrambe le seguenti opzioni.

- **Medie e deviazioni standard.** Visualizzate per ciascuna variabile. Viene indicato anche il numero di casi con valori non mancanti. I valori mancanti vengono gestiti variabile per variabile indipendentemente dall'impostazione corrispondente.
- **Prodotti degli scarti e delle covarianze.** Viene visualizzato per ciascuna coppia di variabili. La deviazione del prodotto incrociato è equivalente alla somma dei prodotti delle variabili corrette per la media. È il numeratore del coefficiente di correlazione di Pearson. La covarianza è una misura non standardizzata della relazione tra due variabili, equivalente alla deviazione del prodotto incrociato divisa per *N*–1.

Valori mancanti. È possibile scegliere tra le opzioni seguenti:

- **Esclusione pairwise.** I casi con valori mancanti per una o entrambe le variabili di una coppia per un coefficiente di correlazione vengono esclusi dall'analisi. Poiché ciascun coefficiente si basa su tutti i casi con codici validi per quella particolare coppia di variabili, in tutti i calcoli verrà utilizzato il maggior numero di informazioni disponibile. In questo modo è possibile ottenere una serie di coefficienti basati su un numero variabile di casi.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile vengono esclusi da tutte le correlazioni.

Funzioni aggiuntive dei comandi CORRELATIONS e NONPAR CORR

Il linguaggio della sintassi dei comandi consente inoltre di:

- Scrivere una matrice di correlazione per la correlazione di Pearson da utilizzare in luogo dei dati per ottenere altri tipi di analisi, ad esempio l'analisi fattoriale (con il sottocomando `MATRIX`).
- Ottenere correlazioni di ciascuna variabile presente in un elenco con le variabili corrispondenti presenti in un secondo elenco (utilizzando la parola chiave `WITH` con il sottocomando `VARIABLES`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
26

Correlazioni parziali

La procedura Correlazioni parziali consente di calcolare i coefficienti di correlazione parziale che descrivono la relazione lineare tra due variabili controllando gli effetti di una o più variabili aggiuntive. Le correlazioni sono misure di associazione lineare. Due variabili possono essere perfettamente correlate, ma se la relazione non è lineare, un coefficiente di correlazione non rappresenta la statistica più adatta a misurarne l'associazione.

Esempio. Esiste una correlazione tra i fondi stanziati per il sistema sanitario e il tasso di malattie? Sebbene si potrebbe pensare che una relazione di tale tipo sia negativa, da uno studio emerge una correlazione *positiva* significativa: con l'aumentare dei fondi stanziati per il sistema sanitario, il tasso di malattie sembra aumentare. Il controllo della frequenza delle visite ai medici elimina virtualmente la correlazione positiva osservata. I fondi stanziati per il sistema sanitario e il tasso di malattie sembrano avere una correlazione positiva solo perché più pazienti hanno accesso al sistema sanitario quando vengono aumentati i fondi, con un conseguente aumento delle malattie segnalate da medici e ospedali.

Statistiche. Per ogni variabile: numero di casi con valori non mancanti, media e deviazione standard. Matrici di correlazione parziale e di ordine zero, con gradi di libertà e livelli di significatività.

Dati. Utilizzare variabili simmetriche, quantitative.

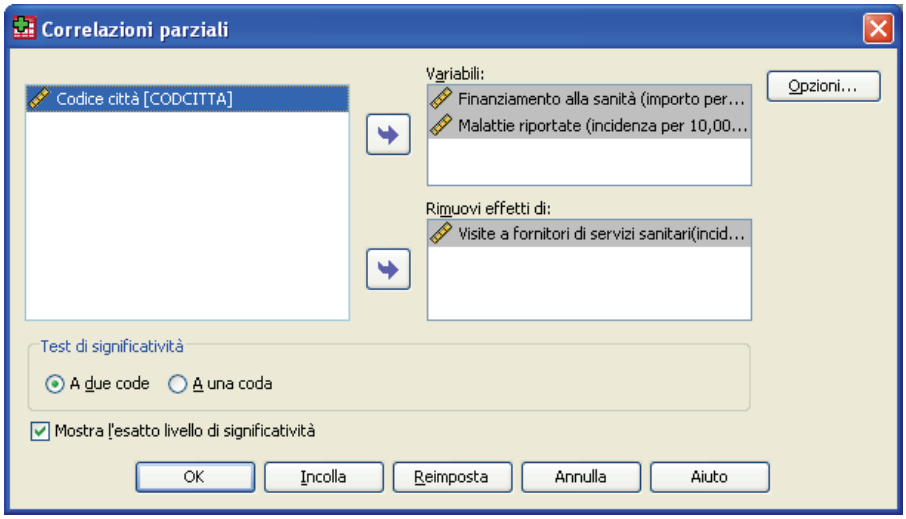
Assunzioni. La procedura Correlazioni parziali si fonda sull'ipotesi che ciascuna coppia di variabili sia bivariata normale.

Per ottenere correlazioni parziali

► Dai menu, scegliere:

Analizza
  Correlazione
    Parziale...

Figura 26-1
Finestra di dialogo Correlazioni parziali



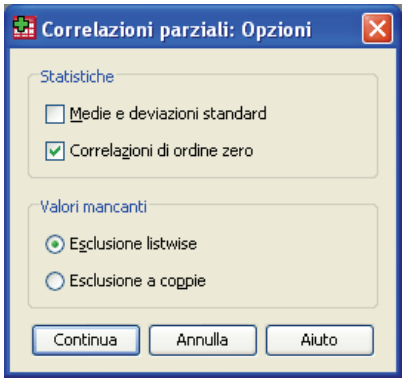
- Selezionare due o più variabili numeriche per cui è necessario calcolare le correlazioni parziali.
- Selezionare una o più variabili numeriche di controllo.

Sono inoltre disponibili le seguenti opzioni:

- **Test di significatività.** È possibile selezionare le probabilità a una coda o a due code. Se si conosce in anticipo la direzione dell'associazione, selezionare A una coda.. In alternativa, selezionare A due code.
- **Mostra l'esatto livello di significatività.** Per impostazione predefinita, vengono indicati la probabilità e i gradi di libertà di ciascun coefficiente di correlazione. Se questa opzione viene deselezionata, i coefficienti significativi al livello 0,05 vengono identificati con un solo asterisco, i coefficienti significativi al livello 0,01 con un doppio asterisco e i gradi di libertà vengono eliminati. Questa impostazione viene applicata alle matrici di correlazione parziale e di ordine zero.

Correlazioni parziali: Opzioni

Figura 26-2
Finestra di dialogo Correlazioni parziali: Opzioni



Statistiche. È possibile scegliere una delle seguenti opzioni o entrambe:

- **Medie e deviazioni standard.** Visualizzate per ciascuna variabile. Viene indicato anche il numero di casi con valori non mancanti.
- **Correlazioni di ordine zero.** Viene visualizzata una matrice di correlazioni semplici tra tutte le variabili, incluse le variabili di controllo.

Valori mancanti. È possibile scegliere una delle seguenti opzioni:

- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile, incluse le variabili di controllo, vengono esclusi da tutti i calcoli.
- **Esclusione pairwise.** Per il calcolo delle correlazioni di ordine zero su cui si basano le correlazioni parziali, i casi con valori mancanti per una o entrambe le variabili di una coppia non verranno utilizzati. L'eliminazione pairwise consente di utilizzare il massimo numero di dati possibile. Il numero di casi, tuttavia, può differire a seconda del coefficiente. Quando è attiva l'eliminazione pairwise, i gradi di libertà per un particolare coefficiente parziale si basano sul numero minimo di casi utilizzati per il calcolo di una delle correlazioni di ordine zero.

Opzioni aggiuntive del comando PARTIAL CORR

Il linguaggio della sintassi dei comandi consente inoltre di:

- Leggere una matrice di correlazione di ordine zero o scrivere una matrice di correlazione parziale (con il sottocomando `MATRIX`).
- Ottenere correlazioni parziali tra due elenchi di variabili (con la parola chiave `WITH` nel sottocomando `VARIABLES`).
- Ottenere più analisi (con più sottocomandi `VARIABLES`).
- Specificare i valori degli ordini da richiedere (ad esempio sia le correlazioni parziali di primo e secondo ordine) quando sono disponibili due variabili di controllo (con il sottocomando `VARIABLES`).
- Eliminare i coefficienti ridondanti (con il sottocomando `FORMAT`).
- Visualizzare una matrice di correlazioni semplici quando non è possibile calcolare alcuni coefficienti (con il sottocomando `STATISTICS`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
27

Distanze

Questa procedura consente di calcolare una grande varietà di statistiche in base alle similarità o alle dissimilarità (distanze), sia considerando coppie di variabili, sia considerando coppie di casi. Tali misure di distanza o similarità possono poi essere utilizzate insieme ad altre procedure quali analisi fattoriale, analisi cluster o scaling multidimensionale per l'analisi di insiemi complessi di dati.

Esempio. È possibile misurare le similarità tra coppie di automobili in base a caratteristiche specifiche, quali dimensione del motore, MPG e potenza. Grazie al calcolo delle similarità, è possibile stabilire quali automobili sono simili e quali sono differenti tra loro. Per un'analisi più formale, si può scegliere di applicare alle similarità l'analisi cluster gerarchica o lo scaling multidimensionale che consentono di esplorare la struttura sottostante.

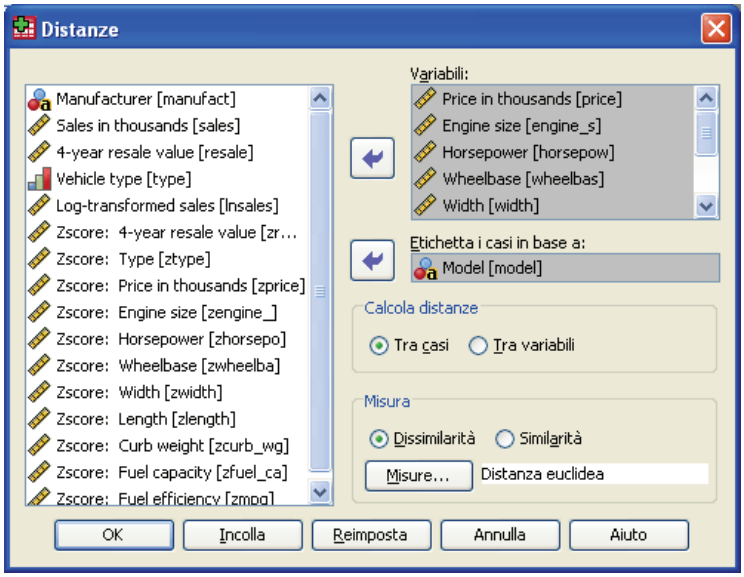
Statistiche. Le misure di dissimilarità (distanza) disponibili per i dati di intervallo sono: distanza euclidea, distanza euclidea quadratica, Chebychev, City-block, Minkowski, personalizzata. Per i dati di conteggio, le misure disponibili sono chi-quadrato o phi-quadrato. Per i dati binari, infine, le misure disponibili sono: distanza euclidea, distanza euclidea quadratica, differenza di dimensione, differenza di modello, varianza, forma, Lance e Williams. Le misure di similarità disponibili per i dati di intervallo sono la correlazione di Pearson o il coseno. Per i dati binari sono invece disponibili le seguenti misure: Russel e Rao, corrispondenza semplice, Jaccard, Dice, Rogers e Tanimoto, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Kulczynski 1, Kulczynski 2, Sokal e Sneath 4, Hamann, Lambda, *D* di Anderberg, *Y* di Yule, *Q* di Yule, Ochiai, Sokal e Sneath 5, correlazione phi a 4 punti o dispersione.

Per ottenere matrici delle distanze

- Dai menu, scegliere:

Analizza
  Correlazione
    Distanze...

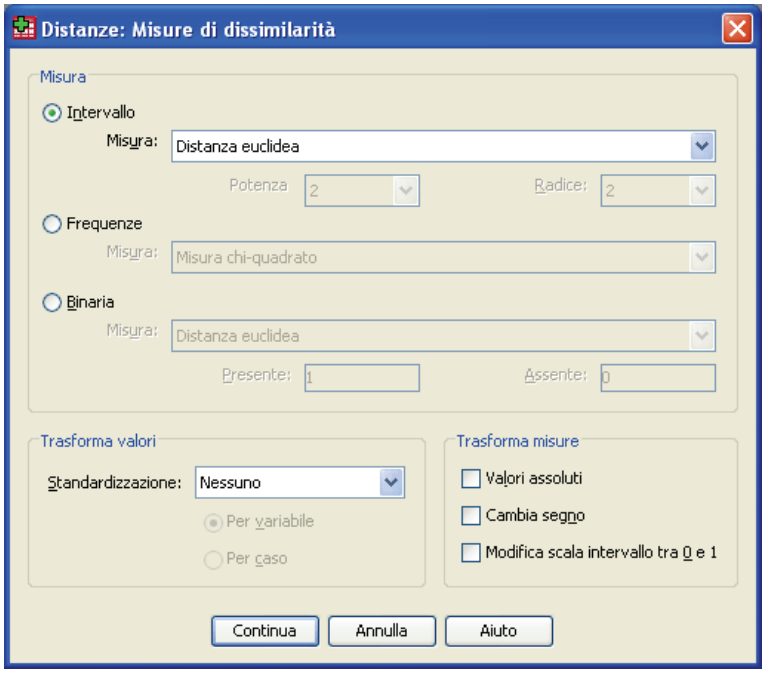
Figura 27-1
Finestra di dialogo Distanze



- Selezionare una variabile numerica per calcolare la distanza tra casi e almeno due variabili numeriche per calcolare la distanza tra variabili.
- Scegliere un'alternativa nel gruppo Calcola distanze per calcolare le distanze tra casi o tra variabili.

Distanze: Misure di dissimilarità

Figura 27-2
Finestra di dialogo Distanze: Misure di dissimilarità



Dal gruppo Misura selezionare l'alternativa corrispondente al tipo di dati desiderato (intervallo, conteggio o binari), quindi scegliere dall'elenco a discesa una delle misure corrispondenti a quel tipo di dati. Le misure disponibili per tipo di dati sono le seguenti:

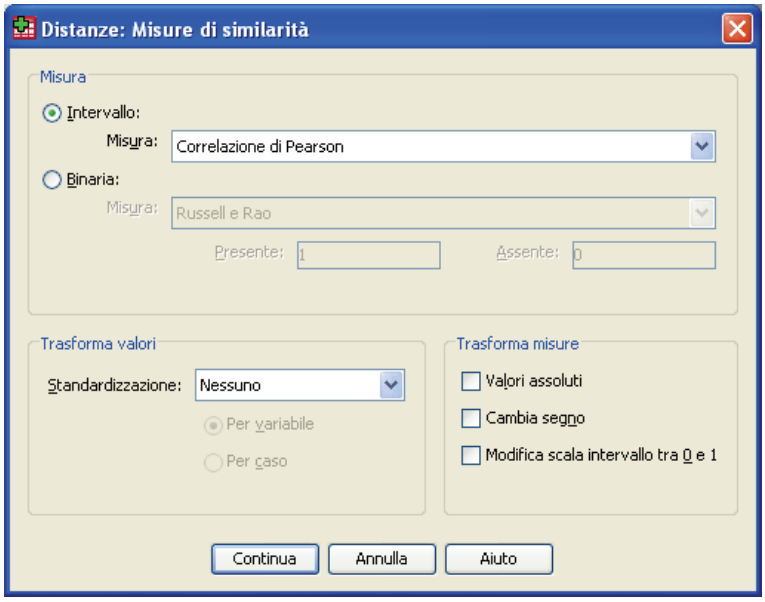
- **Dati per intervallo.** Distanza euclidea, distanza euclidea quadratica, Chebychev, City-block, Minkowski o personalizzata.
- **Conteggio dati.** Misura chi-quadrato e misura phi-quadrato.
- **Dati binari.** Distanza euclidea, distanza euclidea quadratica, differenza di dimensione, differenza di modello, varianza, forma o di Lance e Williams. (Inserire i valori Presente e Assente per specificare i due valori significativi, tutti gli altri valori verranno ignorati dalle distanze).

Il gruppo Trasforma valori consente di standardizzare i valori per casi o valori *prima* di calcolare le similarità. Tali trasformazioni non sono applicabili ai dati binari. I metodi di standardizzazione disponibili sono: punteggi *z*, intervallo da –1 a 1, intervallo da 0 a 1, ampiezza massima di 1, media di 1 o deviazione standard di 1.

Il gruppo Trasforma misure consente di trasformare i valori generati dalla misura della distanza. Questi verranno applicati dopo il calcolo della misura di distanza. Le alternative disponibili sono: Valori assoluti, Cambia segno e Riscala all'intervallo tra 0 e 1.

Distanze: Misure di similarità

Figura 27-3
Finestra di dialogo Distanze: Misure di similarità



Selezionare l'alternativa corrispondente al tipo di dati desiderato (intervallo o binari) dal gruppo Misura, quindi scegliere una delle misure corrispondenti a quel tipo di dati dall'elenco a discesa. Le misure disponibili per tipo di dati sono le seguenti:

- **Dati per intervallo.** Correlazione Pearson o coseno.
- **Dati binari.** Russel e Rao, corrispondenza semplice, Jaccard, Dice, Rogers e Tanimoto, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Kulczynski 1, Kulczynski 2, Sokal e Sneath 4, Hamann, Lambda, *D* di Anderberg, *Y* di Yule, *Q* di Yule, Ochiai, Sokal e Sneath 5, correlazione phi a 4 punti o dispersione. (Inserire i valori Presente e Assente per specificare i due valori significativi, tutti gli altri valori verranno ignorati dalle distanze).

Il gruppo Trasforma valori consente di standardizzare i valori per casi o valori prima di calcolare le similarità. Tali trasformazioni non sono applicabili ai dati binari. I metodi di standardizzazione disponibili sono: punteggi *z*, intervallo da –1 a 1, intervallo da 0 a 1, ampiezza massima di 1, media di 1 e deviazione standard di 1.

Il gruppo Trasforma misure consente di trasformare i valori generati dalla misura della distanza. Questi verranno applicati dopo il calcolo della misura di distanza. Le alternative disponibili sono: Valori assoluti, Cambia segno e Riscala all'intervallo tra 0 e 1.

Opzioni aggiuntive del comando PROXIMITIES

La procedura Distanze usa la sintassi del comando `PROXIMITIES`. Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare un numero intero come potenza per la misura della distanza di Minkowski.
- Specificare qualsiasi intero come potenza e radice per la misura personalizzata delle distanze.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
28

Regressione lineare

La regressione lineare consente di stimare i coefficienti dell'equazione lineare, includendo una o più variabili indipendenti, che prevedono al meglio il valore della variabile dipendente. Ad esempio, è possibile tentare di prevedere le vendite annuali di un rappresentante (la variabile dipendente) in base a variabili indipendenti quali l'età, gli studi e gli anni di esperienza lavorativa.

Esempio. Il numero di partite vinte da una squadra di basket in una stagione è correlato al numero medio di punti effettuati dalla squadra per partita? Un grafico a dispersione indica che queste variabili sono correlate in modo lineare. Il numero di partite vinte e il numero medio di punti effettuati dalla squadra avversaria sono anch'essi correlati in modo lineare. Queste variabili hanno una relazione negativa. Al crescere del numero delle partite vinte, diminuisce il numero medio di punti effettuati dall'avversario. Con la regressione lineare, è possibile modellare la relazione di queste variabili. È possibile utilizzare un modello valido per stimare quante partite vinceranno le squadre.

Statistiche. Per ogni variabile: numero di casi validi, media e deviazione standard. Per ogni modello: coefficienti di regressione, matrice di correlazione, correlazioni di ordine zero e parziali, *R* multipli, $R^{quadrato}$, R^2 corretto, variazioni $R^{quadrato}$, errore standard della stima, tabella di analisi della varianza, valori attesi e residui. Inoltre, intervalli di confidenza al 95% per ogni coefficiente di regressione, matrice di varianza-covarianza, fattore d'inflazione della varianza, tolleranza, test di Durbin-Watson, misure di distanza (Mahalanobis, Cook e valori di influenza), DiffBeta, DiffAdatt, intervalli di stima e diagnostiche per casi. Grafici: grafici a dispersione, grafici parziali, istogrammi e grafici di probabilità normale.

Dati. Le variabili dipendenti ed indipendenti devono essere quantitative. È necessario che le variabili categoriali, come la religione, l'età o la regione di residenza, siano ricodificate come variabili binarie (fittizie) o altri tipi di variabili di contrasto.

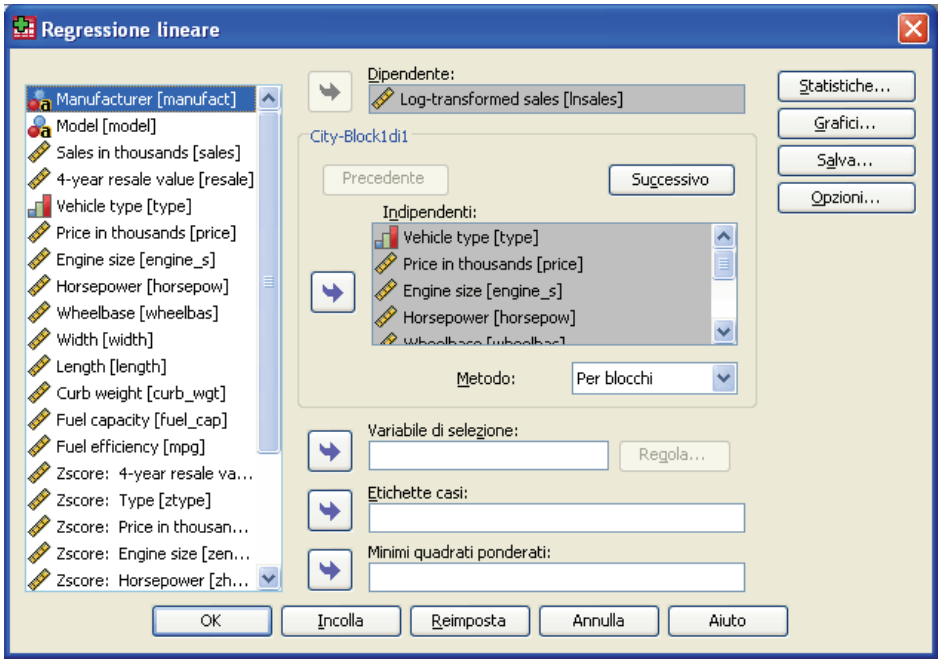
Assunzioni. Per ciascun valore della variabile indipendente, la distribuzione della variabile dipendente deve essere normale. La varianza della distribuzione della variabile dipendente deve essere costante per tutti i valori della variabile indipendente. La relazione tra la variabile dipendente e ogni variabile indipendente deve essere lineare e tutte le osservazioni devono essere indipendenti.

Per ottenere un'analisi della regressione lineare

▶ Dai menu, scegliere:

Analizza
  Regression
    Lineare...

Figura 28-1
Finestra di dialogo Regressione lineare



- Nella finestra di dialogo Regressione lineare, selezionare una variabile dipendente numerica.
- Selezionare una o più variabili indipendenti numeriche.

Se lo si desidera, è possibile:

- Raggruppare le variabili indipendenti in blocchi e specificare metodi di inserimento differenti per sottogruppi di variabili diversi.
- Scegliere una variabile di selezione per limitare l'analisi a un sottoinsieme di casi con valori particolari per questa variabile.
- Selezionare una variabile di casi per l'identificazione di punti nei grafici.
- Selezionare una variabile di ponderazione per WLS per un'analisi dei minimi quadrati ponderati.

Minimi quadrati ponderati (WLS). Consente di ottenere un modello dei minimi quadrati ponderati. I dati vengono ponderati in base al reciproco della loro varianza. Le osservazioni con varianza elevata hanno un peso minore nell'analisi rispetto a quelle con varianza ridotta. Casi con valore 0, negativo o mancante per la variabile di ponderazione saranno esclusi dall'analisi.

Metodi di selezione della variabile di regressione lineare

La selezione del metodo consente di specificare come vengono inserite nell'analisi le variabili indipendenti. Utilizzando diversi metodi, è possibile creare molteplici modelli di regressione dallo stesso insieme di variabili.

- **Per blocchi (Regressione).** Una procedura per la selezione delle variabili nella quale tutte le variabili di un blocco sono inserite in un unico passo.
- **Per passi.** Ad ogni passo viene inserita la variabile indipendente non presente nell'equazione che ha la più bassa probabilità di F, se tale probabilità è sufficientemente piccola. Le variabili già presenti nell'equazione di regressione vengono rimosse se la loro probabilità di F diviene sufficientemente elevata. Il metodo termina quando nessuna variabile rispetta il criterio di inserimento o quello di rimozione.
- **Rimozione.** Una procedura per la selezione di variabili in cui tutte le variabili di un blocco sono rimosse in un solo passo.
- **Eliminazione all'indietro.** Una procedura di selezione di variabili nella quale tutte le variabili vengono inserite nell'equazione e poi rimosse sequenzialmente. La variabile con la più bassa correlazione parziale rispetto alla variabile dipendente viene considerata la prima da rimuovere e viene rimossa se soddisfa il criterio di eliminazione. Dopo la rimozione della prima variabile, la variabile con la più bassa correlazione parziale tra quelle rimaste nell'equazione viene considerata come la prossima da eliminare. La procedura termina quando nell'equazione nessuna variabile soddisfa il criterio di rimozione.
- **Selezione in avanti.** Una procedura di selezione delle variabili nella quale le variabili vengono inserite in modo sequenziale all'interno del modello. La prima variabile da inserire nell'equazione è quella con la più elevata correlazione positiva o negativa con la variabile dipendente. Questa variabile viene inserita nell'equazione solo se soddisfa il criterio di inserimento. Se è stata inserita la prima variabile, viene considerata come successiva la variabile indipendente non presente nell'equazione che ha la più elevata correlazione parziale. La procedura termina quando non ci sono più variabili che soddisfano il criterio di inserimento.

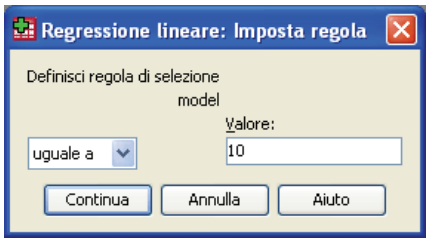
I valori di significatività dell'output si basano sull'adattamento di un singolo modello. Pertanto, i valori di significatività in genere non sono validi quando viene utilizzato un metodo stepwise (per passi, avanti o indietro).

Tutte le variabili devono soddisfare il criterio di tolleranza per essere inserite nell'equazione, indipendentemente dal metodo di inserimento specificato. Il livello di tolleranza predefinito è 0,0001. Una variabile non viene inserita se può far sì che la tolleranza di un'altra variabile già nel modello non rientri nel criterio di tolleranza già stabilito.

Tutte le variabili indipendenti selezionate vengono aggiunte a un solo modello di regressione. È tuttavia possibile specificare diversi metodi di inserimento per diversi sottoinsiemi di variabili. Ad esempio, è possibile inserire un blocco di variabili nel modello di regressione utilizzando la selezione per passi e un secondo blocco utilizzando la selezione in avanti. Per aggiungere un secondo blocco di variabili a un modello di regressione, fare clic su Avanti.

Regressione lineare: Imposta regola

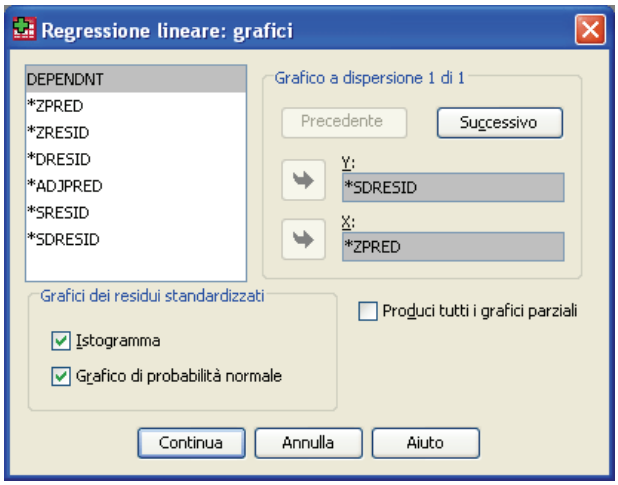
Figura 28-2
Finestra di dialogo Regressione lineare: Imposta regola



Nell'analisi verranno inseriti i casi definiti dalla regola di selezione impostata. Ad esempio, se si seleziona la variabile Uguale a e si immette 5 per il valore, nell'analisi verranno inclusi solo i casi in cui la variabile selezionata ha un valore uguale a 5. È inoltre possibile specificare un valore stringa.

Regressione lineare: grafici

Figura 28-3
Finestra di dialogo Regressione lineare: grafici



I grafici possono facilitare la validazione delle ipotesi di normalità, linearità e uguaglianza delle varianze. I grafici sono inoltre utili per la rilevazione di valori anomali, osservazioni insolite e casi di influenza. Dopo averli salvati come nuove variabili, sarà possibile utilizzare valori attesi, residui e altre diagnostiche disponibili nell'Editor dei dati per la creazione di grafici con le variabili indipendenti. Sono disponibili i seguenti grafici:

Grafici a dispersione. Consentono di rappresentare due elementi qualsiasi tra i seguenti: la variabile dipendente, i valori attesi standardizzati, i residui standardizzati, i residui cancellati, i valori attesi corretti, i residui studentizzati o i residui cancellati studentizzati. Rappresentare nel grafico i residui standardizzati e i valori attesi standardizzati per verificare la linearità e l'uguaglianza delle varianze.

Elenco di variabili sorgente. Elenca la variabile dipendente (DEPENDNT) e le seguenti variabili (valori stimati e residui): valori stimati standardizzati (*ZPRED), residui standardizzati (*ZRESID), residui cancellati (*DRESID), valori stimati corretti (*ADJPRED), residui studentizzati (*SRESID), residui cancellati studentizzati (*SDRESID).

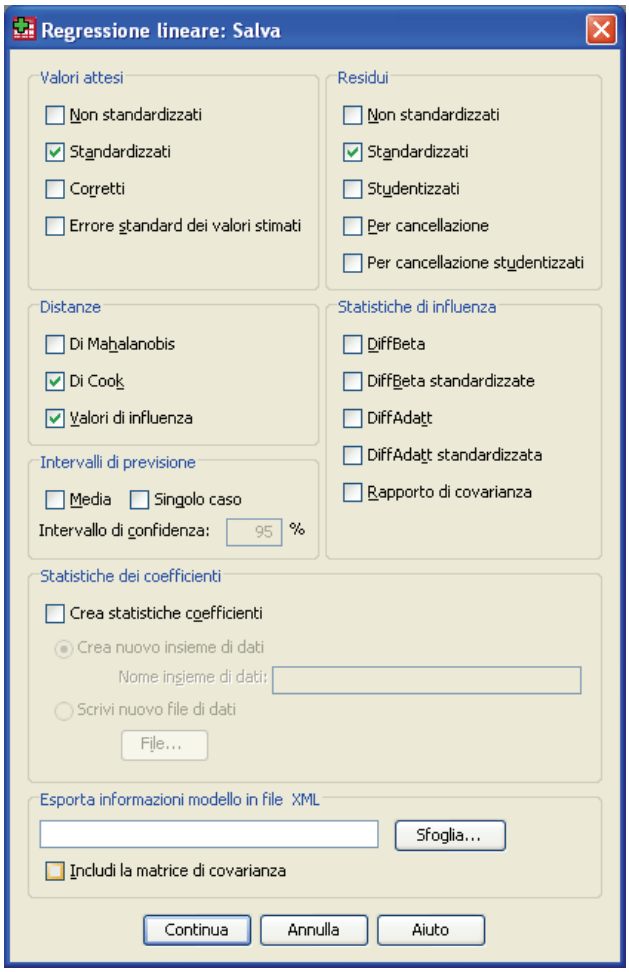
Produci tutti i grafici parziali. Consente di visualizzare i grafici a dispersione dei residui di ogni variabile indipendente e i residui della variabile dipendente quando entrambe le variabili sono regresse separatamente dal resto delle variabili indipendenti. Per la creazione di un grafico parziale è necessario che nell'equazione siano rappresentate almeno due variabili indipendenti.

Grafici dei residui standardizzati . Consentono di ottenere istogrammi dei residui standardizzati e grafici di probabilità normale confrontando la distribuzione dei residui standardizzati e la distribuzione normale.

Se vengono richiesti grafici, verranno visualizzate statistiche riassuntive per i valori attesi standardizzati e per i residui standardizzati (*ZPRED e *ZRESID).

Regressione lineare: Per salvare nuove variabili

Figura 28-4
Finestra di dialogo Regressione lineare: Salva



È possibile salvare valori attesi, residui e altre statistiche utili per la diagnostica. Ogni selezione aggiunge una o più nuove variabili al file dati attivo.

Valori attesi. I valori previsti dal modello di regressione per ogni caso.

- **Non standardizzati.** I valori risultanti dal modello per la variabile dipendente e per ciascun caso.
- **Standardizzati.** Una trasformazione di ciascun valore atteso nella sua forma standardizzata. ovvero il valore atteso medio viene sottratto dal valore atteso e la differenza viene divisa per la deviazione standard dei valori attesi. I valori attesi standardizzati hanno media 0 e deviazione standard 1.
- **Corretti.** Il valore atteso per un caso quando quel caso è escluso dal calcolo dei coefficienti di regressione.
- **Errore standard dei valori stimati.** L'errore standard della media per i valori stimati. Una stima della deviazione standard del valore medio della variabile dipendente per i casi che hanno gli stessi valori delle variabili indipendenti.

Distanze. Misure per l'identificazione dei casi con combinazioni di valori insolite per le variabili indipendenti e dei casi che possono avere un notevole peso sul modello di regressione.

- **Di Mahalanobis.** Una misura della distanza di un caso dalla media di tutti i casi per le variabili indipendenti. Un'elevata distanza di Mahalanobis indica che un caso include valori estremi per una o più variabili indipendenti.
- **Di Cook.** Una misura di quanto cambierebbero i residui di tutti i casi se un particolare caso fosse escluso dal calcolo dei coefficienti di regressione. Valori alti indicano che l'esclusione di un caso dal calcolo dei coefficienti di regressione ne modificherebbe sostanzialmente il valore.
- **Valori di influenza.** Una misura dell'influenza di un dato sull'adattamento della regressione. L'influenza centrata varia da 0 (nessuna influenza sull'adattamento) a (N-1)/N.

Intervalli di previsione. I limiti superiore ed inferiore per gli intervalli di previsione singoli e medi.

- **Media.** Limiti inferiore e superiore (due variabili) per l'intervallo di stima della risposta media prevista.
- **Singolo caso.** Limiti inferiore e superiore (due variabili) per l'intervallo di stima della variabile dipendente per un singolo caso.
- **Intervallo di confidenza.** Immettere un valore compreso fra 1 e 99,99 per specificare l'intervallo di confidenza per i due intervalli di stima. Per disporre di questa opzione è necessario aver selezionato Media o Individuale. I valori tipici dell'intervallo di confidenza sono 90, 95 e 99.

Residui. Il valore effettivo della variabile dipendente meno il valore atteso dall'equazione di regressione.

- **Non standardizzati.** La differenza tra un valore osservato e il valore stimato dal modello.
- **Standardizzati.** Il residuo diviso per una stima della deviazione standard. Il residuo standardizzato, conosciuto anche come residuo di Pearson, ha media 0 e deviazione standard 1.
- **Studentizzati.** Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, a seconda della distanza tra i valori assunti per questo caso dalle variabili indipendenti e le medie delle variabili indipendenti.
- **Per cancellazione.** Il residuo per un caso se quel caso venisse escluso dal calcolo dei coefficienti di regressione. È la differenza tra il valore della variabile dipendente e il valore stimato corretto.
- **Per cancellazione studentizzati.** Il residuo cancellato per un caso diviso per il suo errore standard. La differenza tra un residuo cancellato studentizzato e il suo corrispondente residuo studentizzato indica quanta differenza produca l'eliminazione di un caso sulla stima del medesimo.

Statistiche di influenza. La modifica nei coefficienti di regressione (DiffBeta) e nei valori attesi (DiffAdatt) che risultano dall'esclusione di un particolare caso. I valori DiffBeta e DiffAdatt standardizzati sono anche disponibili con il rapporto di covarianza.

- **Differenza in beta.** Variazione del coefficiente di regressione quando un caso particolare viene eliminato dall'analisi. Viene calcolato un valore per ogni termine del modello, incluso il termine costante.

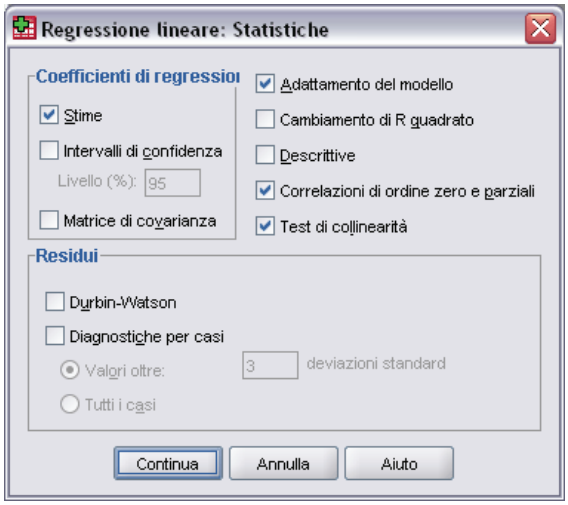
- **DiffBeta standardizzate.** La differenza standardizzata nel valore beta. La variazione di un coefficiente di regressione quando un caso viene rimosso dall'analisi. Possono essere esaminati i casi con valore assoluto superiore a 2 diviso per la radice quadrata di N, dove N è il numero dei casi. Viene calcolato un valore per ogni termine del modello, incluso il termine costante.
- **DiffAdatt.** La differenza nel valore adattato è il cambiamento del valore previsto quando un caso viene escluso dall'analisi.
- **DiffAdatt standardizzata.** La differenza standardizzata nel valore adattato. La variazione del valore stimato quando un caso viene rimosso dall'analisi. Si possono esaminare valori standardizzati maggiori in valore assoluto a 2 per la radice quadrata di p/N, dove p è il numero di parametri del modello e N è il numero di casi.
- **Rapporto di covarianza.** Rapporto fra il determinante della matrice di covarianza con un caso escluso dal calcolo dei coefficienti di regressione ed il determinante con tutti i casi inclusi. Se il rapporto è vicino a 1, il caso non altera significativamente la matrice di covarianza.

Statistiche dei coefficienti. Salva i coefficienti di regressione in un file di dati. I file di dati possono anche essere riutilizzati nella stessa sessione, ma non vengono salvati come file a meno che siano stati salvati come tali alla fine della sessione. I nomi degli insiemi di dati devono essere conformi alle regole dei nomi delle variabili. Per ulteriori informazioni, vedere Nomi delle variabili in Capitolo 5 a pag. 79.

Esporta informazioni modello in file XML. Le stime dei parametri e, se si desidera, le relative covarianze vengono esportati nel file specificato in formato XML (PMML). SmartScore e SPSS Statistics Server (un prodotto separato) possono utilizzare questo file del modello per applicare le informazioni del modello ad altri file dati a scopo di assegnazione dei punteggi.

Regressione lineare: Statistiche

Figura 28-5
Finestra di dialogo Statistiche



Sono disponibili le seguenti statistiche:

Coefficienti di regressione. Stime consente di visualizzare il coefficiente di regressione *B*, l'errore standard di *B*, il coefficiente beta standardizzato, il valore *t* per *B*e il livello di significatività a due code di *t*. Intervalli di confidenza visualizza gli intervalli di confidenza con il livello di confidenza specificato per ciascun coefficiente di regressione o matrice di covarianza. Matrice di covarianza consente di visualizzare una matrice di varianza-covarianza dei coefficienti di regressione con le covarianze esterne alla diagonale e le varianze sulla diagonale. Viene inoltre visualizzata una matrice di correlazione.

Adattamento del modello. Vengono elencate le variabili inserite ed eliminate dal modello e vengono visualizzate le seguenti statistiche di bontà dell'adattamento: *R* multipli, $R^{quadrato}$ e $R^{quadrato}$ corretto, errore standard della stima e una tabella di analisi della varianza.

Cambiamento di R quadrato. La variazione nella statistica $R^{quadrato}$ che viene prodotta aggiungendo o eliminando una variabile indipendente. Se la variazione di $R^{quadrato}$ associata ad una variabile è elevata, ciò significa che la variabile è un valido stimolatore della variabile dipendente.

Descrittive. Fornisce il numero di casi validi, la media e la deviazione standard per ogni variabile nell'analisi. Vengono inoltre visualizzati una matrice di correlazione con un livello di significatività a una coda e il numero di casi per ogni correlazione.

Correlazione parziale. Ciò che rimane della correlazione fra due variabili, dopo aver rimosso gli effetti della loro reciproca correlazione con altre variabili. Ad esempio, la correlazione fra la variabile dipendente e una variabile indipendente dopo aver rimosso da entrambe gli effetti lineari delle altre variabili indipendenti incluse nel modello.

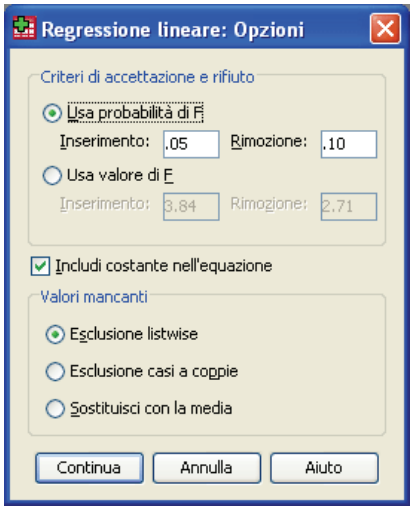
Correlazione parziale. La correlazione fra la variabile dipendente e una variabile indipendente, dopo aver rimosso dalla variabile indipendente gli effetti lineari della correlazione con altre variabili indipendenti. Correlata alla variazione di R-quadrato quando una variabile viene aggiunta a un'equazione. A volte detta correlazione semiparziale.

Diagnostiche di collinearità. La collinearità (o multicollinearità) è la situazione in cui una delle variabili indipendenti è una funzione lineare di altre variabili indipendenti. Consente di visualizzare gli autovalori della matrice dei prodotti incrociati scalata e non centrata, l'indice di collinearità e le proporzioni della decomposizione della varianza con i fattori di inflazione della varianza (VIF) e le tolleranze per le singole variabili.

Residui. Consente di visualizzare il test di Durbin-Watson per la correlazione seriale dei residui e la diagnostica per i casi che corrispondono al criterio di selezione (valori anomali superiori a *n* deviazioni standard).

Regressione lineare: Opzioni

Figura 28-6
Finestra di dialogo Regressione lineare: Opzioni



Sono disponibili le seguenti opzioni:

Criteri di accettazione e rifiuto. Queste opzioni vengono utilizzate quando si specifica il metodo di selezione delle variabili in avanti, indietro o per passi. È possibile inserire o eliminare le variabili dal modello in base alla significatività (probabilità) del valore *F* o allo stesso valore *F*.

- **Usa probabilità di F.** La variabile viene inserita nel modello se il livello di significatività del relativo valore di F è minore di quello di inserimento. La variabile viene altresì rimossa se il livello di significatività è maggiore di quello di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi e Inserimento deve essere minore di Rimozione. Alzando il valore di inserimento e/o abbassando quello di rimozione si allentano i vincoli di inclusione delle variabili nel modello.
- **Usa valore di F.** La variabile viene inserita nel modello se il relativo valore F è maggiore di quello di inserimento. La variabile viene altresì rimossa se il relativo valore F è minore di quello di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi e Inserimento deve essere maggiore di Rimozione Abbassando il valore di inserimento e/o alzando quello di rimozione si allentano i vincoli di inclusione delle variabili nel modello.

Includi termine costante nell'equazione. Per impostazione predefinita, il modello di regressione include un termine costante. Se l'opzione è deselezionata, viene forzato il passaggio della curva di regressione per l'origine, il che avviene raramente. Alcuni risultati di una curva di regressione che passa per l'origine non sono confrontabili con i risultati della regressione che include una costante. Ad esempio $R^{quadrato}$ non può essere interpretato nel modo usuale.

Valori mancanti. È possibile scegliere tra le opzioni seguenti:

- **Esclusione listwise.** Sono inclusi nell'analisi solo i casi con valori validi per tutte le variabili.

- **Esclusione pairwise.** Per calcolare il coefficiente di correlazione su cui si basa l'analisi della regressione vengono utilizzati i casi con dati completi per la coppia di variabili correlate. I gradi di libertà sono basati su *N* pairwise minimo.
- **Sostituisci con la media.** Per i calcoli vengono utilizzati tutti i casi e la media della variabile viene sostituita alle osservazioni mancanti.

Opzioni aggiuntive del comando REGRESSION

Il linguaggio della sintassi dei comandi consente inoltre di:

- Scrivere una matrice di correlazione o leggere una matrice anziché i dati grezzi per ottenere l'analisi di regressione (con il sottocomando `MATRIX`).
- Specificare i livelli di tolleranza (tramite il sottocomando `CRITERIA`).
- Ottenere più modelli per variabili dipendenti uguali o diverse (con i sottocomandi `METHOD` e `DEPENDENT`).
- Ottenere statistiche aggiuntive (con i sottocomandi `DESCRIPTIVES` e `STATISTICS`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
29

Regressione ordinale

La procedura Regressione ordinale consente di definire la dipendenza di una risposta ordinale politomica in un insieme di stimatori, che possono essere fattori o covariate. La struttura della regressione ordinale è basata sulla metodologia di McCullagh (1980, 1998) e la procedura è denominata `PLUM` nella sintassi.

L'analisi della regressione lineare standard comporta la riduzione al minimo della somma dei quadrati delle differenze tra una variabile di risposta (dipendente) e una combinazione ponderata di variabili stimatore (indipendenti). I coefficienti stimati riflettono il modo in cui le modifiche dei predittori influiscono sulla risposta. Si presume che la risposta sia numerica e ciò significa che le modifiche al livello di risposta sono uguali nell'intervallo della risposta. Ad esempio, la differenza di altezza tra una persona alta 150 cm e una persona alta 140 cm è di 10 cm ed equivale alla differenza di altezza tra una persona alta 210 cm e una persona alta 200 cm. Queste relazioni non sono necessariamente valide per le variabili ordinali, nelle quali la scelta e il numero delle categorie di risposta possono essere arbitrarie.

Esempio. È possibile utilizzare la regressione ordinale per studiare la reazione dei pazienti a un dosaggio medicinale. Le possibili reazioni possono essere classificate come *nessuna*, *lieve*, *moderata* o *grave*. La differenza tra una reazione lieve e una moderata è molto difficile o impossibile da quantificare ed è basata sulla percezione. In generale, la differenza tra una risposta lieve e una moderata può essere maggiore o minore della differenza tra una risposta moderata e una grave.

Statistiche e grafici. Frequenze osservate e attese e frequenze cumulate, residui di Pearson per le frequenze e frequenze cumulate, probabilità osservate e attese, probabilità osservate e probabilità cumulate attese per ciascuna categoria di risposta in base al modello covariato, correlazione asintotica e matrici di covarianza delle stime dei parametri, chi-quadrato di Pearson e chi-quadrato del rapporto di verosimiglianza, statistiche sulla bontà di adattamento, cronologia delle iterazioni, test dell'ipotesi di linee parallele, stime dei parametri, errori standard, intervalli di confidenza e statistiche R^2 di Cox e Snell, di Nagelkerke e di McFadden.

Dati. Si presume che la variabile dipendente sia ordinale e può essere numerica o stringa. L'ordinamento è determinato dalla disposizione dei valori della variabile dipendente in ordine crescente, in cui il valore più basso definisce la prima categoria. Si presume che le variabili fattore siano categoriali, mentre le covariate devono essere numeriche. Si noti che l'utilizzo di più covariate continue in genere comporta la creazione di una tabella di probabilità di cella di dimensioni molto grandi.

Assunzioni. È consentita una sola variabile di risposta, che deve essere specificata. Inoltre, per ciascun modello distinto di valori nelle variabili indipendenti, si presume che le risposte siano variabili multinomiali indipendenti.

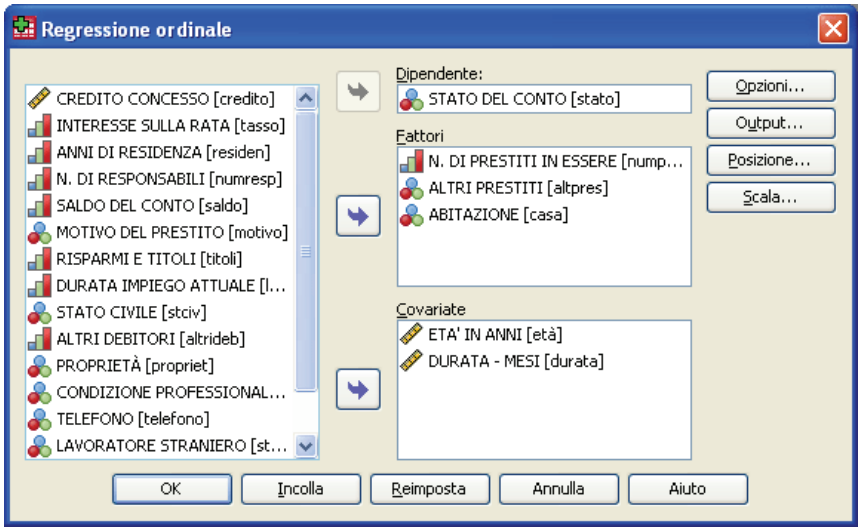
Procedure correlate. La regressione logistica nominale utilizza modelli simili per le variabili dipendenti nominali.

Per ottenere una regressione ordinale

▶ Dai menu, scegliere:

Analizza
  Regressione
    Ordinale...

Figura 29-1
Finestra di dialogo Regressione ordinale



▶ Selezionare una variabile dipendente.

▶ Fare clic su OK.

Regressione ordinale: Opzioni

Nella finestra di dialogo Opzioni è possibile correggere i parametri utilizzati nell'algoritmo di stima iterativo, scegliere un livello di confidenza per le stime dei parametri e selezionare una funzione di collegamento.

Figura 29-2
Finestra di dialogo Regressione ordinale: Opzioni

Iterazioni. È possibile personalizzare l'algoritmo iterativo.

- **Massimo numero di iterazioni.** Specificare un intero non negativo. Se si specifica 0, la procedura restituisce le stime iniziali.
- **Max dimezzamenti.** Specificare un intero positivo.
- **Convergenza log-verosimiglianza.** L'algoritmo si interrompe se il cambiamento assoluto o relativo della verosimiglianza è inferiore a questo valore. Il criterio non viene utilizzato se si specifica 0.
- **Convergenza parametri.** L'algoritmo si interrompe se il cambiamento assoluto o relativo in ciascuna delle stime dei parametri è inferiore a questo valore. Il criterio non viene utilizzato se si specifica 0.

Intervallo di confidenza. Specificare un valore maggiore o uguale a 0 e minore di 100.

Delta. Valore aggiunto alle frequenze zero di cella. Specificare un valore non negativo inferiore a 1.

Tolleranza della singolarità. Utilizzata per controllare gli stimatori a dipendenza elevata. Selezionare un valore dall'elenco delle opzioni.

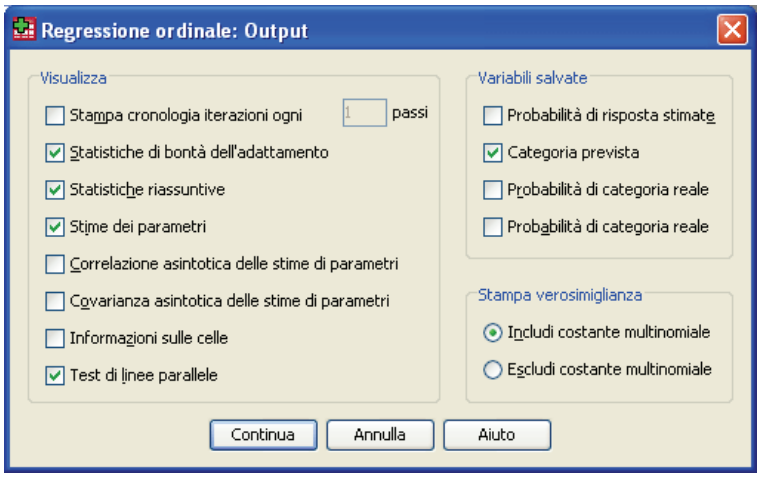
Funzione di collegamento. La funzione Collegamento è la trasformazione delle probabilità cumulate che permette di stimare il modello. Sono disponibili le cinque funzioni di collegamento descritte nella tabella che segue.

Funzione	Formato	Applicazione tipica
Logit	$\log(\xi / (1-\xi))$	Categorie distribuite uniformemente
Log-log complementare	$\log(-\log(1-\xi))$	Categorie alte più probabili
Log-log negativo	$-\log(-\log(\xi))$	Categorie basse più probabili
Probit	$\Phi^{-1}(\xi)$	La variabile latente è normalmente distribuita
Cauchit (funzione Cauchy inversa)	$\tan(\pi(\xi-0.5))$	La variabile latente ha molti valori estremi

Regressione ordinale: Output

Nella finestra di dialogo Output è possibile creare tabelle da visualizzare nel Viewer e salvare variabili nel file di lavoro.

Figura 29-3
Finestra di dialogo Regressione ordinale: Output



Visualizza. Consente di creare tabelle per:

- **Stampa cronologia iteraz. ogni n passi.** Stampa le stime della verosimiglianza e dei parametri per la frequenza di iterazioni di stampa specificata. La prima e l'ultima iterazione vengono sempre stampate.
- **Statistiche sulla bontà dell'adattamento.** Statistiche chi-quadrato di Pearson e del rapporto di verosimiglianza. Vengono elaborate in base alla classificazione specificata nell'elenco di variabili.
- **Statistiche riassuntive.** Statistiche R^2 di Cox e Snell, di Nagelkerke e di McFadden.
- **Stime dei parametri.** Stime dei parametri, errori standard e intervalli di confidenza.
- **Correlazione asintotica delle stime di parametri.** Matrice delle correlazioni delle stime dei parametri.
- **Covarianza asintotica delle stime di parametri.** Matrice delle covarianze delle stime dei parametri.
- **Informazioni sulle celle.** Frequenze osservate e attese e frequenze cumulate, residui di Pearson per le frequenze e le frequenze cumulate, probabilità osservate e attese, probabilità cumulate osservate e attese per ciascuna categoria di risposta in base al modello covariato. Si noti che per i modelli che includono più modelli covariati, ad esempio i modelli con covariate continue, questa opzione può creare una tabella di dimensioni molto grandi e difficile da gestire.
- **Test di linee parallele.** Test di verifica dell'ipotesi di equivalenza dei parametri di posizione nei livelli della variabile dipendente. È disponibile per il modello di sola posizione.

Variabili salvate. Salva le seguenti variabili nel file di lavoro:

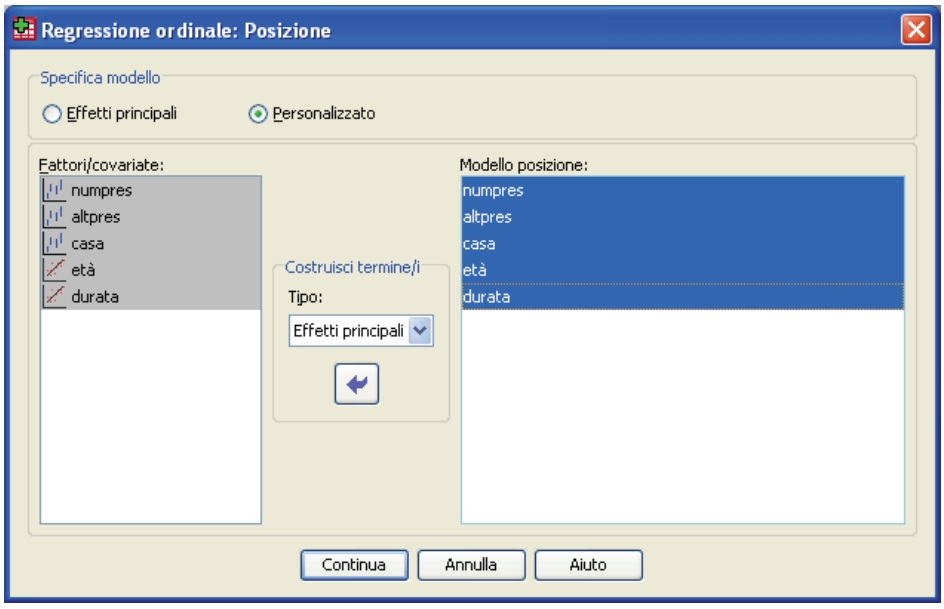
- **Probabilità di risposta stimate.** Probabilità stimate dal modello per la classificazione di un modello fattore o covariato nelle categorie di risposta. Il numero di probabilità corrisponde al numero di categorie di risposta.
- **Categoria prevista.** La categoria di risposta che ha la maggiore probabilità stimata per un modello fattore o covariato.
- **Probabilità di categoria prevista.** Probabilità stimata di classificazione di un modello fattore o covariato nella categoria prevista. La probabilità corrisponde inoltre al massimo di probabilità stimate del modello fattore o covariato.
- **Probabilità di categoria reale.** Probabilità stimata di classificazione di un modello fattore o covariato nella categoria reale.

Stampa verosimiglianza. Controlla la visualizzazione della verosimiglianza. L'inclusione della costante multinomiale consente di ottenere il valore completo della verosimiglianza. Per confrontare i risultati dei prodotti che non includono la costante, si può scegliere di escluderla.

Regressione ordinale: Posizione

Nella finestra di dialogo Posizione è possibile specificare il modello di posizione per l'analisi.

Figura 29-4
Finestra di dialogo Regressione ordinale: Posizione



Specifica modello. Un modello a effetti principali include gli effetti principali di covariate e fattori, ma non gli effetti di interazione. È possibile creare un modello personalizzato per specificare i sottoinsiemi di interazioni di fattori o di interazioni di covariate.

Fattori/covariate. I fattori e le covariate sono elencati.

Modello posizione. Il modello dipende dagli effetti principali e di interazione selezionati.

Costruisci termini

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti 2-vie. Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti 3-vie. Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

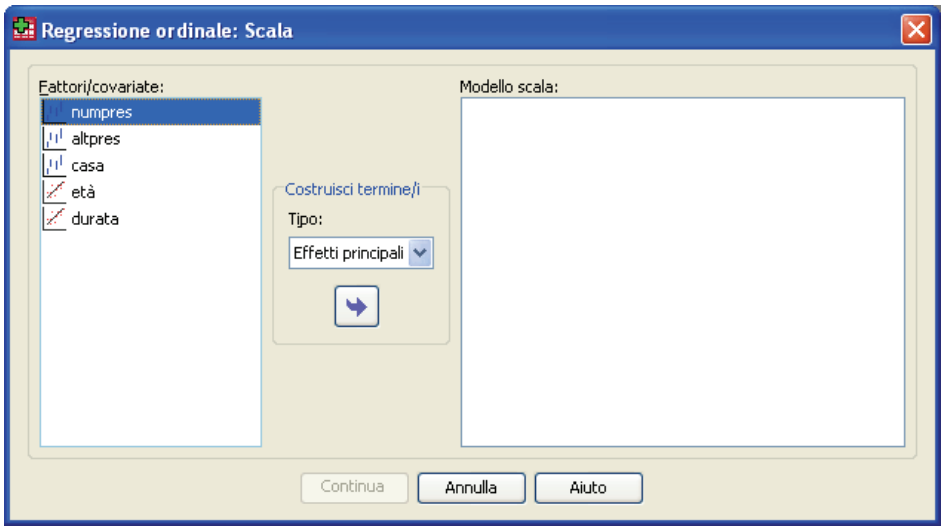
Tutti 4-vie. Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti 5-vie. Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Regressione ordinale: Scala

Nella finestra di dialogo Scala è possibile specificare il modello di scala per l'analisi.

Figura 29-5
Finestra di dialogo Regressione ordinale: Scala



Fattori/covariate. I fattori e le covariate sono elencati.

Modello scala. Il modello dipende dagli effetti principali e di interazione selezionati.

Costruisci termini

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti 2-vie. Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti 3-vie. Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti 4-vie. Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti 5-vie. Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Opzioni aggiuntive del comando PLUM

Per personalizzare la procedura Regressione ordinale è possibile incollare le impostazioni selezionate in una finestra di sintassi e quindi modificare la sintassi del comando `PLUM` così ottenuta. Il linguaggio della sintassi dei comandi consente inoltre di:

- Creare test di ipotesi personalizzati specificando ipotesi nulle come combinazioni lineari dei parametri.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
30

Stima di curve

La procedura Stima di curve produce le statistiche di regressione per la stima di curve e i grafici correlati per 11 diversi modelli di regressione per la stima di curve. Per ciascuna variabile dipendente verrà creato un modello distinto. È inoltre possibile salvare come nuove variabili i valori attesi, i residui e gli intervalli di stima.

Esempio. Un provider di servizi Internet deve tener traccia della percentuale di traffico e-mail infettato da virus sulla propria rete nell'arco di un periodo di tempo specifico. Il grafico di dispersione indica che la relazione non è lineare. È necessario adattare ai dati un modello quadratico o cubico e controllare la validità delle assunzioni e la bontà di adattamento del modello.

Statistiche. Per ciascun modello: coefficienti di regressione, *R* multiplo, $R^{quadrato}$, $R^{quadrato}$ corretto, errore standard della stima, tabella di analisi della varianza, valori attesi, residui ed intervalli di stima. Modelli: lineare, logaritmico, inverso, quadratico, cubico, di potenza, composto, curva S, logistico, di crescita ed esponenziale.

Dati. Le variabili dipendenti ed indipendenti devono essere quantitative. Se si seleziona Tempo come variabile indipenente dal file di dati attivo anziché selezionare una variabile, la procedura Stima di curve genera una variabile tempo se il periodo di tempo tra i casi è uniforme. Se si seleziona Tempo, la variabile dipendente deve essere una serie storica. L'analisi di serie storiche richiede nel file di dati una struttura in cui ciascun caso (riga) rappresenta una serie di osservazioni eseguite a orari diversi e il periodo di tempo fra i casi è uniforme.

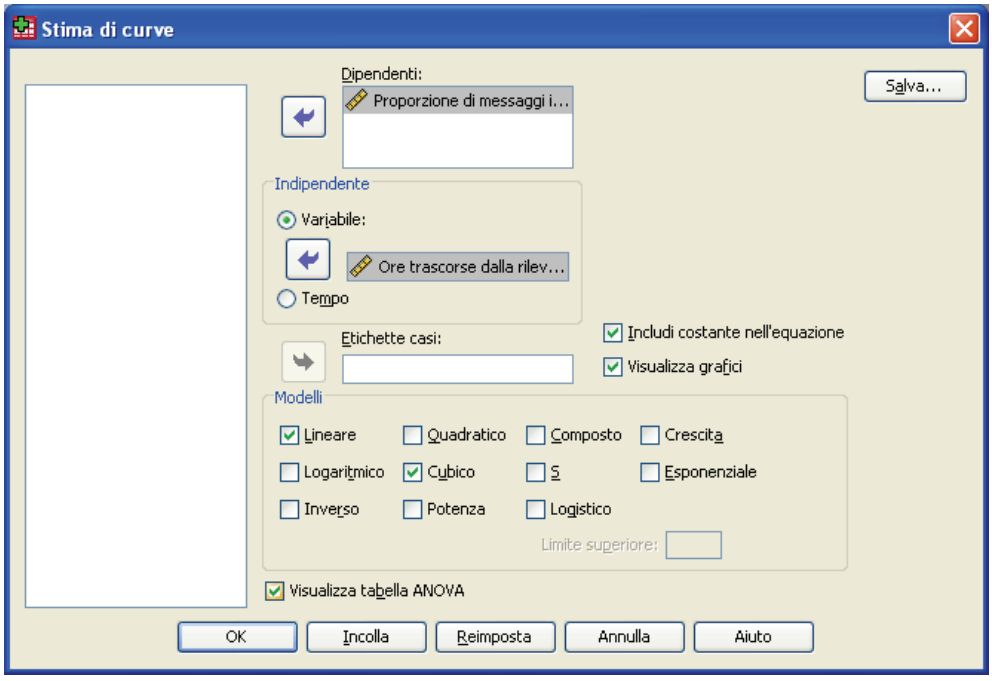
Assunzioni. Valutare i dati graficamente per determinare il tipo di relazione esistente tra variabili dipendenti e indipendenti (lineare, esponenziale, ecc.). I residui di un buon modello devono essere normali e distribuiti casualmente. Se si utilizza un modello lineare, devono essere soddisfatte le seguenti ipotesi. Per ciascun valore della variabile indipendente, la distribuzione della variabile dipendente deve essere normale. La varianza della distribuzione della variabile dipendente deve essere costante per tutti i valori della variabile indipendente. La relazione tra la variabile dipendente e la variabile indipendente deve essere lineare e tutte le osservazioni devono essere indipendenti.

Per ottenere una stima di curve

▶ Dai menu, scegliere:

Analizza
  Regressione
    Stima di curve...

Figura 30-1
Finestra di dialogo Stima di curve



- Selezionare una o più variabili dipendenti. Per ciascuna variabile dipendente verrà creato un modello distinto.
- Selezionare una variabile indipendente (una variabile nel file dati attivo oppure Tempo).
- Oppure:
  - Selezionare una variabile per etichettare casi nei grafici a dispersione. Per ciascun punto del grafico a dispersione, è possibile utilizzare lo strumento di selezione dei punti per visualizzare il valore della variabile Etichetta di caso.
  - Fare clic su Salva per salvare i valori attesi, i residui e gli intervalli di stima come nuove variabili.

Sono inoltre disponibili le seguenti opzioni:

- **Includi termine costante nell'equazione.** Consente di valutare un termine costante nell'equazione di regressione. La costante viene inclusa per impostazione predefinita.
- **Visualizza grafici.** Consente di tracciare i valori della variabile dipendente e ciascun modello selezionato in base alla variabile indipendente. Viene prodotto un grafico per ogni variabile dipendente.
- **Visualizza tabella ANOVA.** Consente di visualizzare una tabella di analisi della varianza per ciascun modello selezionato.

Stima di curve: Modelli

È possibile scegliere uno o più modelli di regressione per la stima di curve. Per determinare il modello da utilizzare, tracciare i dati in un grafico. Se le variabili appaiono legate da una relazione lineare, utilizzare un modello di regressione lineare semplice. Se la relazione tra le variabili non è lineare, provare a trasformare i dati. Se la trasformazione non risulta utile, può essere necessario utilizzare un modello più complesso. Visualizzare i dati in un grafico a dispersione; se il grafico è simile a una funzione matematica nota, adattare i dati a quel tipo di modello. Se, ad esempio, i dati sono simili a una funzione esponenziale, utilizzare il modello esponenziale.

Lineare. Modello la cui equazione è Y=b0+(b1*t). I valori della serie vengono rappresentati come una funzione lineare del tempo.

Logaritmico. Modello la cui equazione è Y = b0 + (b1 * ln(t)).

Inverso. Modello la cui equazione è Y = b0 + (b1 / t).

Quadratico. Modello la cui equazione è: Y=b0+(b1*t)+(b2*t**2). Il modello quadratico può essere usato per modellare una serie che "decolla" o una serie che si smorza rapidamente.

Cubico. Modello definito dall'equazione: Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3).

Potenza. Modello la cui equazione è Y = b0 * (t**b1) oppure ln(Y) = ln(b0) + (b1 * ln(t)).

Composto. Modello la cui equazione è Y = b0*(b1**t) oppure ln(Y) = ln(b0)+(ln(b1)*t).

Curva S. Modello la cui equazione è Y = e**(b0 + (b1/t)) oppure ln(Y) = b0 + (b1/t).

Logistica. Modello la cui equazione è Y=1/(1/u+(b0*(b1**t))) oppure ln(1/Y-1/u)=ln(b0+(ln(b1)*t)) dove u è il limite superiore. Per utilizzare l'equazione di regressione, specificare il valore limite superiore dopo aver selezionato Logistico. Tale valore deve essere un intero positivo maggiore del valore più alto assunto dalla variabile dipendente.

Crescita. Modello la cui equazione è Y = e**(b0 + (b1 * t)) o ln(Y) = b0 + (b1 * t).

Esponenziale. Modello la cui equazione è Y = b0 * (e**(b1 * t)) o ln(Y) = ln(b0) + (b1 * t).

Stima di curve: Salva

Figura 30-2
Finestra di dialogo Stima di curve: Salva

Salva Variabili. Per ciascun modello selezionato è possibile salvare i valori attesi, i residui (valori osservati della variabile dipendente meno il valore atteso del modello) e gli intervalli di stima (limite superiore e inferiore). I nomi delle nuove variabili e le etichette descrittive vengono visualizzati in una tabella nell'output.

Periodo di previsione. Se come variabile indipendente si seleziona Tempo anzichè una variabile nel file dati attivo, è possibile specificare un periodo di previsione al termine delle serie storiche. È possibile scegliere una delle seguenti opzioni:

- **Prevedi dal periodo di stima fino all'ultimo caso.** Consente di prevedere i valori per tutti i casi del file in base ai casi inclusi nel periodo di stima. Il periodo di stima, visualizzato in fondo alla finestra di dialogo, viene definito nella sottofinestra di dialogo dell'opzione Seleziona casi del menu Dati. Se non è stato definito alcun periodo di stima, i valori verranno previsti in base a tutti i casi.
- **Prevedi fino a.** Consente di prevedere i valori fino alla data, all'ora o al numero di osservazione specificato in base ai casi inclusi nel periodo di stima. Questa funzione può essere usata per prevedere valori futuri nelle serie storiche. Le variabili data definite specificano quali caselle di testo è possibile usare per specificare la fine di un periodo di previsione. Se non vengono definite variabili di data, è possibile specificare l'ultimo numero di osservazione (caso).

Per creare variabili di data, utilizzare l'opzione Definisci date, disponibile nel menu Dati.

Capitolo
31
Regressione parziale minimi quadrati

La procedura Regressione parziale minimi quadrati consente di stimare i modelli di regressione parziale dei minimi quadrati (PLS, nota anche come "proiezione della struttura latente"). PLS è una tecnica predittiva che rappresenta un'alternativa alla regressione dei minimi quadrati ordinari (OLS), alla correlazione canonica o ai modelli di equazioni strutturali e si rivela particolarmente utile quando le variabili predittore sono strettamente correlate o quando il numero dei predittori supera il numero dei casi.

PLS combina le funzioni dell'analisi dei componenti principali e della regressione multipla. Consente innanzitutto di estrarre un insieme di fattori latenti che forniscono la maggiore quantità di informazioni possibile sulla covarianza tra le variabili dipendenti e indipendenti. Quindi, un passo di regressione consente di prevedere i valori delle variabili dipendenti mediante la decomposizione delle variabili indipendenti.

Disponibilità. PLS è un comando di estensione che per la sua esecuzione richiede che nell'apposito sistema sia installato il modulo di estensione Python. È necessario installare separatamente il modulo di estensione PLS e il programma di installazione può essere scaricato dall'indirizzo Web *http://www.spss.com/devcentral* .

Nota: Il modulo di estensione PLS dipende dal software Python. SPSS Inc. non è il titolare né il licenziatario del software Python. Tutti gli utenti Python devono accettare i termini del contratto di licenza Python disponibile al sito Web di Python. SPSS Inc. non avanza alcuna valutazione sulla qualità del programma Python. SPSS Inc. declina qualsiasi responsabilità associata all'uso del programma Python.

Tabelle. La proporzione della varianza spiegata (per fattore latente), i pesi fattoriali latenti, i carichi fattoriali latenti, l'importanza della variabile indipendente nella proiezione (VIP) e le stime dei parametri di regressione (per variabile dipendente) vengono tutti generati per impostazione predefinita.

Grafici. L'importanza della variabile nella proiezione (VIP), i punteggi fattoriali, i pesi fattoriali per i primi tre fattori latenti e la distanza dal modello vengono tutti generati dalla scheda Opzioni.

Livello di misurazione. Le variabili dipendenti e indipendenti (predittore) possono essere di scala, nominali o ordinali. La procedura presume che il livello di misurazione appropriato sia stato assegnato a tutte le variabili, sebbene sia possibile modificare temporaneamente il livello di misurazione di una variabile facendo clic con il pulsante destro del mouse sulla variabile nell'elenco delle variabili sorgente e scegliendo un livello di misurazione dal menu di scelta rapida. Le variabili categoriali (nominali o ordinali) vengono trattate in maniera equivalente dalla procedura.

Codifiche variabili categoriali. La procedura ricodifica temporaneamente le variabili dipendenti categoriali utilizzando le codifiche one-of-c*c* per la durata della procedura. Se esistono categorie *c* di una variabile, la variabile viene archiviata come vettori *c*, con la prima categoria indicata (1,0,...,0), la categoria successiva (0,1,0,...,0), ..., e la categoria finale (0,0,...,0,1). Le variabili dipendenti categoriali vengono rappresentate utilizzando una codifica fittizia; ovvero, semplicemente omettendo l'indicatore corrispondente alla categoria di riferimento.

Ponderazione. I valori dei pesi, prima di essere utilizzati vengono arrotondati ai numeri interi più vicini. I casi con pesi mancanti o con pesi inferiori a 0,5 non vengono utilizzati nelle analisi.

Valori mancanti. I valori mancanti di sistema e definiti dall'utente vengono considerati come non validi.

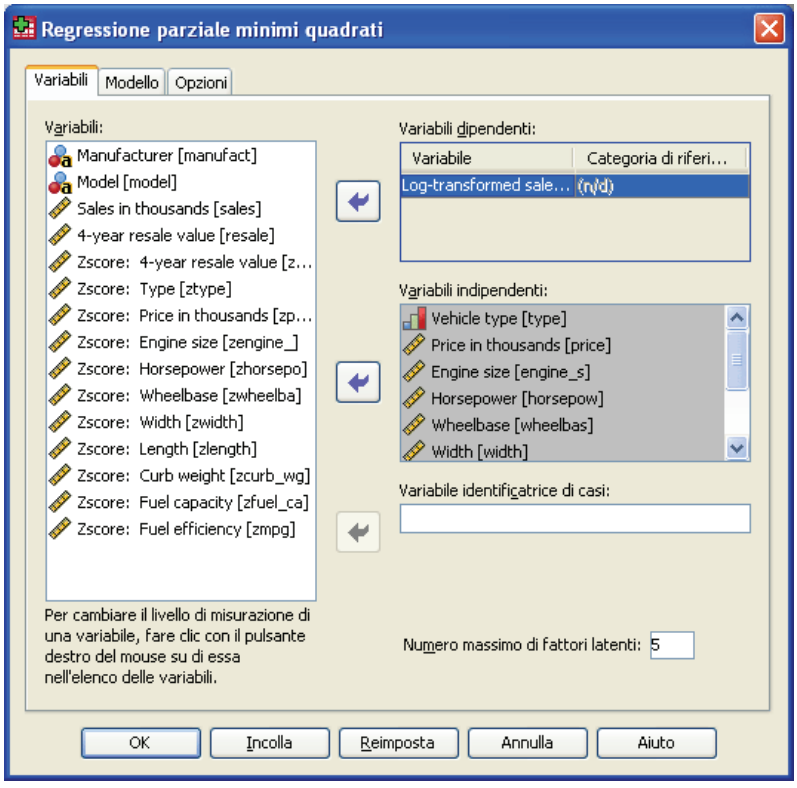
Modifica della scala. Tutte le variabili di modello sono centrate e standardizzate, comprese le variabili indicatore che rappresentano le variabili categoriali.

Per ottenere la regressione parziale dei minimi quadrati

Dai menu, scegliere:

Analizza
  Regression
    Minimi quadrati parziali...

Figura 31-1
Scheda Variabili della finestra Regressione parziale minimi quadrati



► Selezionare almeno una variabile dipendente.

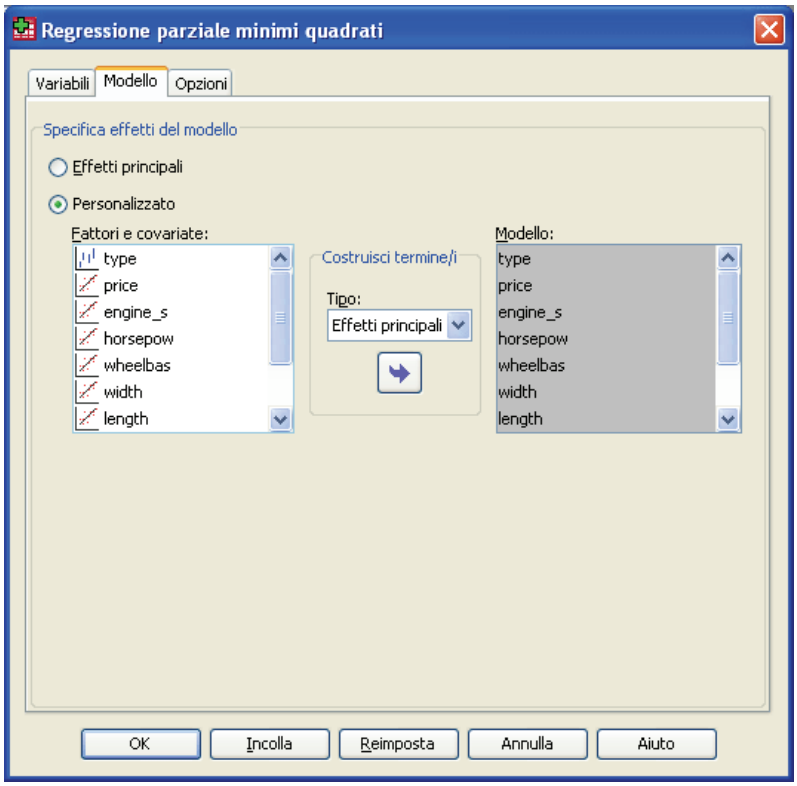
▶ Selezionare almeno una variabile indipendente.

Se lo si desidera, è possibile:

- Specificare una categoria di riferimento per le variabili dipendenti categoriali (nominali o ordinali).
- Specificare una variabile da utilizzare come identificativo univoco per l'output per i casi e gli insiemi di dati salvati.
- Specificare un limite superiore per il numero dei fattori latenti da estrarre.

Modello

Figura 31-2
Scheda Modello della finestra Regressione parziale minimi quadrati



Specifica modello effetti. Un modello di effetti principali include tutti gli effetti principali di covariate e fattori. Selezionare Personalizzato per specificare le interazioni. È necessario indicare tutti i termini da includere nel modello.

Fattori e covariate. I fattori e le covariate sono elencati.

Modello. Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato Personalizzato, è possibile selezionare gli effetti principali e le interazioni desiderate per l'analisi da eseguire.

Costruisci termini

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti 2-vie. Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

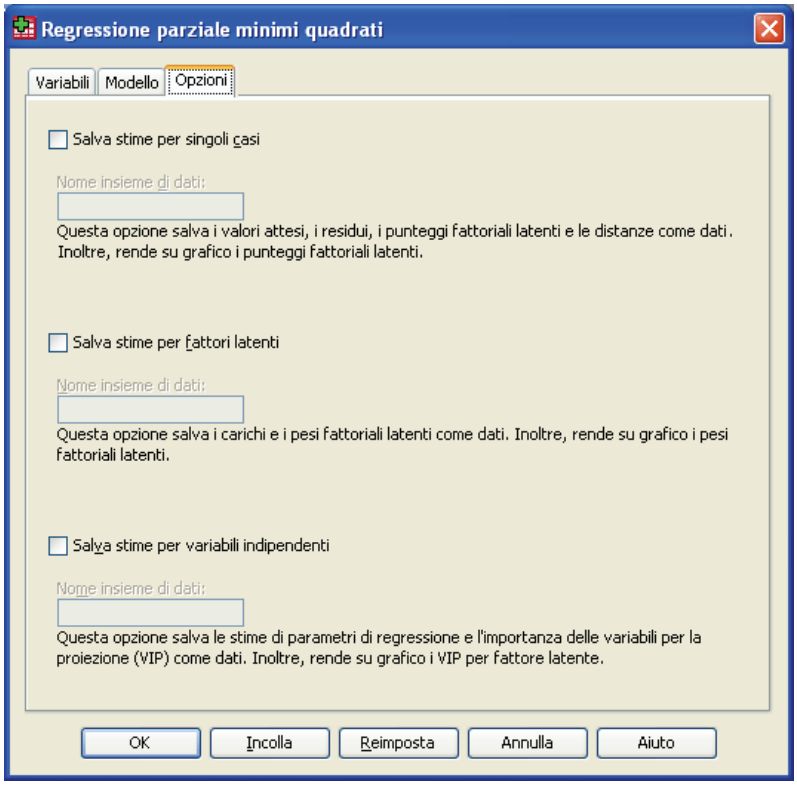
Tutti 3-vie. Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti 4-vie. Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti 5-vie. Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Opzioni

Figura 31-3
Scheda Opzioni della finestra Regressione parziale minimi quadrati


Regressione parziale minimi quadrati
Variabili
Modello
Opzioni
Salva stime per singoli casi
Nome insieme di dati:
Questa opzione salva i valori attesi, i residui, i punteggi fattoriali latenti e le distanze come dati. Inoltre, rende su grafico i punteggi fattoriali latenti.
Salva stime per fattori latenti
Nome insieme di dati:
Questa opzione salva i carichi e i pesi fattoriali latenti come dati. Inoltre, rende su grafico i pesi fattoriali latenti.
Salva stime per variabili indipendenti
Nome insieme di dati:
Questa opzione salva le stime di parametri di regressione e l'importanza delle variabili per la proiezione (VIP) come dati. Inoltre, rende su grafico i VIP per fattore latente.
OK
Incolla
Reimposta
Annulla
Aiuto

La scheda Opzioni consente all'utente di salvare e rappresentare graficamente le stime dei modelli per singoli casi, fattori latenti e predittori.

Per ogni tipo di dati, specificare il nome di un insieme di dati. I nomi degli insiemi di dati devono essere univoci. Se si specifica il nome di un insieme di dati esistente, i suoi contenuti vengono sostituiti; altrimenti, viene creato un nuovo insieme di dati.

- **Salva stime per singoli casi.**Consente di salvare le stime dei modelli caso per caso, ovvero: i valori attesi, i residui, la distanza del modello dei fattori latenti e i punteggi fattoriali latenti. Inoltre, rende su grafico i punteggi fattoriali latenti.
- **Salva stime per fattori latenti.** Consente di salvare i carichi e i pesi fattoriali latenti. Inoltre, rende su grafico i pesi fattoriali latenti.
- **Salva stime per variabili indipendenti.** Consente di salvare le stime dei parametri di regressione e l'importanza delle variabili per la proiezione (VIP). Inoltre, rende su grafico i VIP per fattore latente.

Capitolo
32

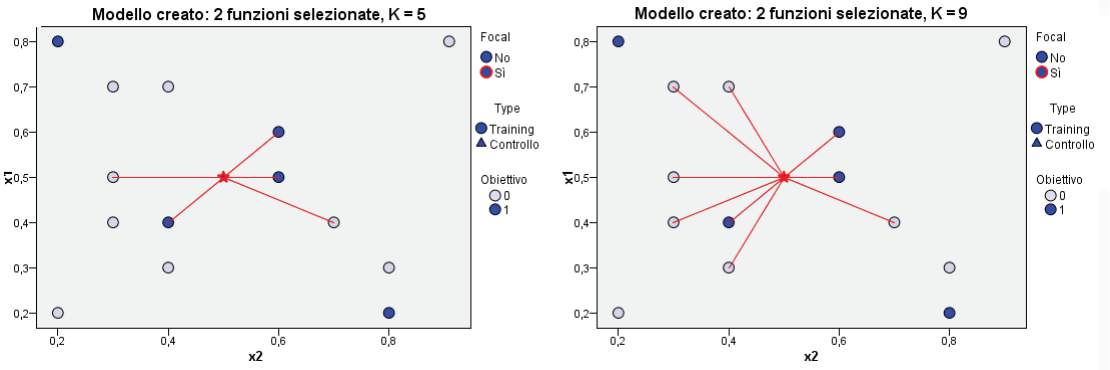
Analisi del vicino più vicino

L'analisi del vicino più vicino è un metodo per la classificazione dei casi basato sulla similarità ad altri casi. Nell'apprendimento automatico, questo metodo è stato sviluppato per riconoscere modelli di dati senza richiedere una corrispondenza esatta con eventuali modelli o casi archiviati. I casi simili sono vicini gli uni agli altri, mentre i casi dissimili sono distanti gli uni dagli altri. Pertanto, la distanza tra due casi rappresenta una misura della loro dissimilarità.

I casi attigui vengono definiti "vicini". Quando viene presentato un nuovo caso, rappresentato dal punto interrogativo, viene calcolata la sua distanza da ogni caso del modello. Le classificazioni dei casi più simili, ovvero i vicini più vicini, vengono registrate e il nuovo caso viene inserito nella categoria contenente il numero più alto di vicini più vicini.

È possibile specificare il numero di vicini più vicini da esaminare; tale valore viene definito *k*. Le immagini mostrano in che modo viene classificato un nuovo caso utilizzando due valori diversi di *k*. Quando $k = 5$, il nuovo caso viene inserito nella categoria *1*, perché la maggioranza dei vicini più vicini appartiene alla categoria *1*. Tuttavia, quando $k = 9$, il nuovo caso viene inserito nella categoria *0*, perché la maggioranza dei vicini più vicini appartiene alla categoria *0*.

Figura 32-1
Effetti della modifica di k sulla classificazione



L'analisi dei vicini più vicini può essere utilizzata anche per calcolare i valori per un obiettivo continuo. In questa situazione viene calcolata una media dei valori dei vicini più vicini sull'obiettivo per ottenere un valore per il nuovo caso.

Obiettivo e funzioni. L'obiettivo e le funzioni possono essere:

- **Nominale.** Una variabile può essere considerata nominale quando i relativi valori rappresentano categorie prive di ordinamento intrinseco, per esempio l'ufficio di una società, la regione, il codice postale e la religione.

- **Ordinale.** Una variabile può essere considerata ordinale quando i relativi valori rappresentano categorie con qualche ordinamento intrinseco, per esempio i gradi di soddisfazione per un servizio, da molto insoddisfatto a molto soddisfatto, i punteggi di atteggiamento corrispondenti a gradi di soddisfazione o fiducia e i punteggi di preferenza.
- **Scala.** Una variabile può essere considerata di scala quando i relativi valori rappresentano categorie ordinate con una metrica significativa, tale che i confronti fra le distanze dei relativi valori siano appropriati. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

  Le variabili nominali e ordinali vengono trattate in modo analogo dall'analisi del vicino più vicino. La procedura presume che il livello di misurazione appropriato sia stato assegnato a ciascuna variabile. Tuttavia, è possibile modificare temporaneamente il livello di misurazione di una variabile facendo clic con il pulsante destro del mouse sulla variabile nell'elenco delle variabili sorgente e scegliendo un livello di misurazione dal menu di scelta rapida. Per modificare in modo permanente il livello di misurazione di una variabile, vedere Livello di misurazione delle variabili.

L'icona accanto a ciascuna variabile nell'elenco delle variabili identifica il livello di misurazione e il tipo di dati.

Livello di misurazione	**Tipo di dati**			
	Numerica	Stringa	Data	Ora
Scala		n/a		
Ordinale				
Nominale				

Codifiche variabili categoriali. La procedura ricodifica temporaneamente le variabili dipendenti e indipendenti categoriali utilizzando le codifiche one-of-*c* per la durata della procedura. Se esistono categorie *c* di una variabile, la variabile viene archiviata come vettori c, con la prima categoria indicata (1,0,...,0), la categoria successiva (0,1,0,...,0), ..., e la categoria finale (0,0,...,0,1).

Questo schema di codifica aumenta la dimensionalità dello spazio delle funzioni. In particolare, il numero totale di dimensioni è pari al numero di predittori di scala più il numero di categorie in tutti i predittori categoriali. Ne consegue che questo schema di codifica può generare un training più lento. Se il training del vicino più vicino sta procedendo molto lentamente, è possibile cercare di ridurre il numero di categorie nei predittori categoriali mediante la combinazione di categorie simili o casi di rilascio con categorie estremamente rare prima dell'esecuzione della procedura.

Tutta la codifica one-of-*c* è basata sui dati di training, anche se viene definito un campione di verifica o di controllo (vedere Partizioni). Pertanto, se il campione di controllo contiene casi con categorie di predittori assenti nei dati di training, non è possibile calcolarne il punteggio. Se il campione di controllo contiene casi con categorie di variabili dipendenti assenti nei dati di training, è invece possibile calcolarne il punteggio.

Modifica della scala. Le funzioni di scala vengono normalizzate per impostazione predefinita. Tale modifica viene eseguita interamente sulla base dei dati di training, anche se viene definito un campione di controllo (vedere Partizioni a pag. 396). Se si specifica una variabile per definire le partizioni, è importante che tali funzioni abbiano distribuzioni simili nei campioni di training e controllo. Utilizzare, ad esempio, la procedura Esplora per esaminare le distribuzioni nelle partizioni.

Ponderazione. La ponderazione viene ignorata da questa procedura.

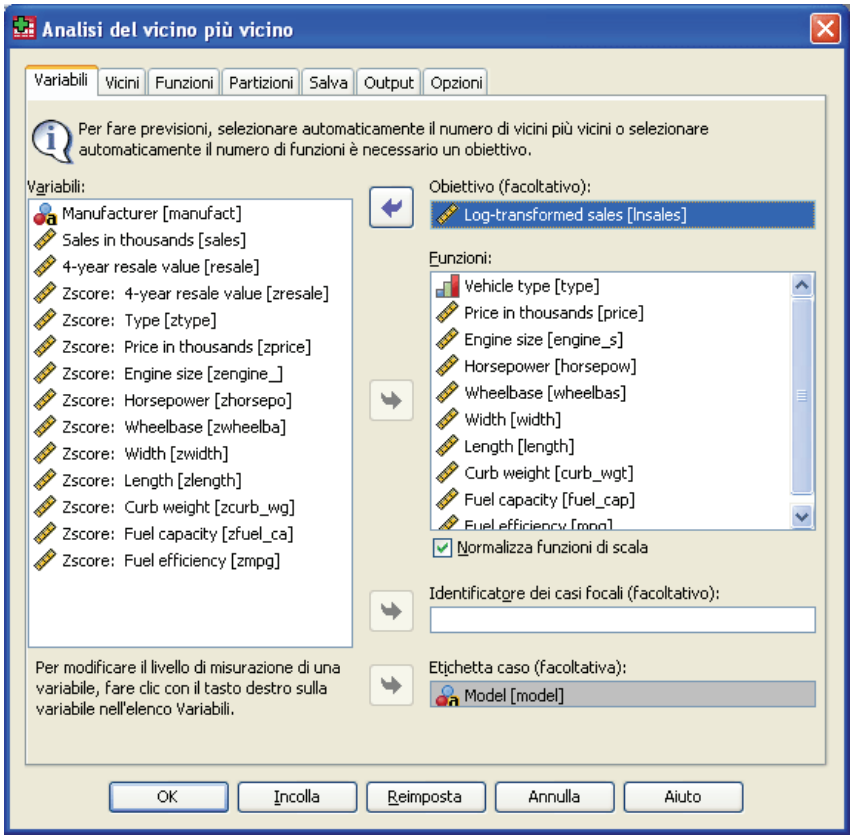
Replica dei risultati. La procedura utilizza la generazione di numeri casuali durante l'assegnazione causale delle partizioni e dei sottocampioni con convalida incrociata. Se si desidera replicare esattamente i risultati ottenuti, oltre a utilizzare le stesse impostazioni per la procedura, è necessario impostare il valore di seme per il generatore di Mersenne Twister (vedere Partizioni a pag. 396) o utilizzare le variabili per definire le partizioni e i sottocampioni con convalida incrociata.

Per ottenere un'analisi del vicino più vicino

Dai menu, scegliere:

Analizza
  Classifica
    Vicino più vicino...

Figura 32-2
Finestra di dialogo Analisi del vicino più vicino, scheda Variabili



- Specificare una o più funzioni, che possono essere considerate variabili o predittori indipendenti in presenza di un obiettivo.

Obiettivo (facoltativo). Se non viene specificato alcun obiettivo (variabile o risposta dipendente), vengono rilevati esclusivamente i *k* vicini più vicini. Non vengono eseguite classificazioni né previsioni.

Normalizza funzioni di scala. Le funzioni normalizzate hanno lo stesso intervallo di valori, il che può migliorare le prestazioni dell'algoritmo di stima. Viene utilizzata la normalizzazione corretta, [2*(x−min)/(max−min)]−1. I valori normalizzati corretti sono compresi tra −1 e 1.

Identificatore dei casi focali (facoltativo). Consente di contrassegnare casi di particolare interesse. Si supponga, ad esempio, che un ricercatore desideri stabilire se i punteggi dei test di un determinato distretto scolastico (il caso focale) sono paragonabili a quelli di distretti scolastici analoghi. Ricorrerà all'analisi del vicino più vicino per individuare i distretti scolastici con le maggiori analogie in termini di uno specifico insieme di funzioni. Procederà quindi al confronto tra i punteggi dei test del distretto scolastico focale e quelli dei vicini più vicini.

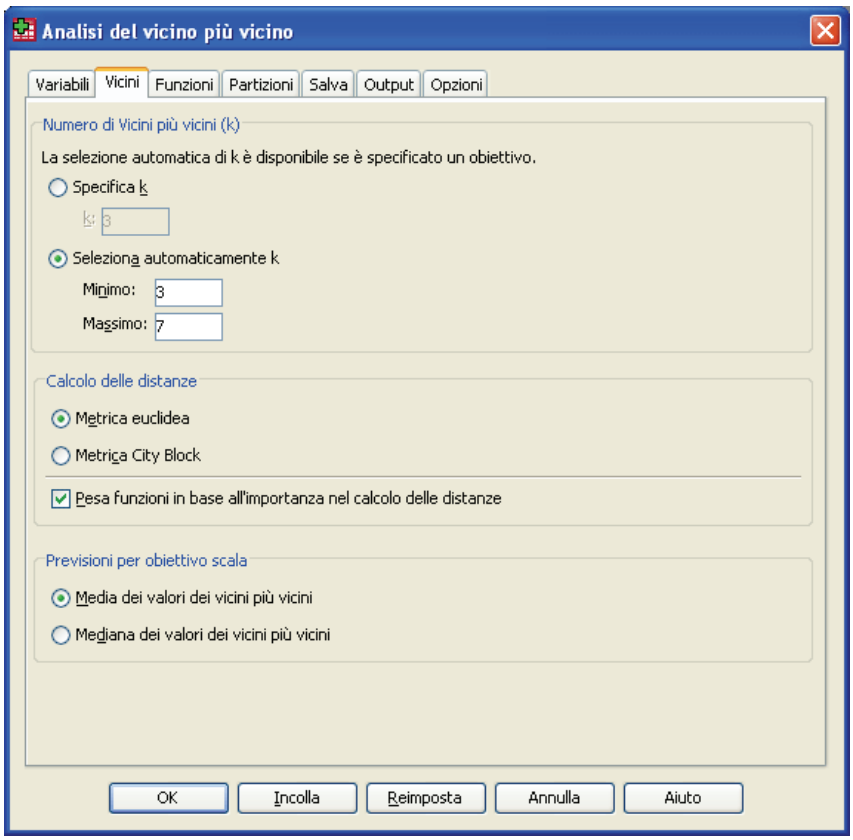
I casi focali possono essere usati anche negli studi clinici per selezionare casi di controllo simili a casi clinici. I casi focali vengono visualizzati nella tabella dei *k* vicini più vicini e delle distanze, nel grafico dello spazio delle funzioni, in quello degli equivalenti e nella mappa dei quadranti. Le informazioni sui casi focali vengono salvate nei file specificati nella scheda Output.

I casi con un valore positivo per la variabile specificata vengono trattati come casi focali. Non è consentito specificare una variabile senza valori positivi.

Etichetta caso (facoltativa). Ai casi vengono applicate etichette utilizzando questi valori nel grafico dello spazio di funzioni, in quello degli equivalenti e nella mappa dei quadranti.

Vicini

Figura 32-3
Finestra di dialogo Analisi del vicino più vicino, scheda Vicini



Numero di Vicini più vicini (k). Specificare il numero di vicini più vicini. L'utilizzo di un numero maggiore di vicini non garantisce necessariamente un modello più preciso.

Se nella tabella Variabili è specificato un obiettivo, in alternativa è possibile indicare un intervallo di valori e lasciare che sia la procedura a scegliere il "miglior" numero di vicini all'interno di tale intervallo. Il metodo utilizzato per stabilire il numero di vicini più vicini dipende dalla necessità o meno della selezione delle funzioni nella scheda Funzioni.

- Se la selezione delle funzioni è attiva, viene eseguita per ciascun valore di *k* nell'intervallo richiesto e viene selezionato il *k* (con il relativo insieme di funzioni) con il tasso di errore più basso (o l'errore più basso della somma dei quadrati se l'obiettivo è una scala).
- Se, invece, la selezione delle funzioni non è attiva, viene utilizzata la convalida incrociata con sottocampioni *V* per selezionare il "miglior" numero di vicini. Per informazioni sul controllo dell'assegnazione dei sottocampioni, vedere la scheda Partizione.

Calcolo delle distanze. Metrica utilizzata per specificare la metrica di distanza per la misurazione della similarità dei casi.

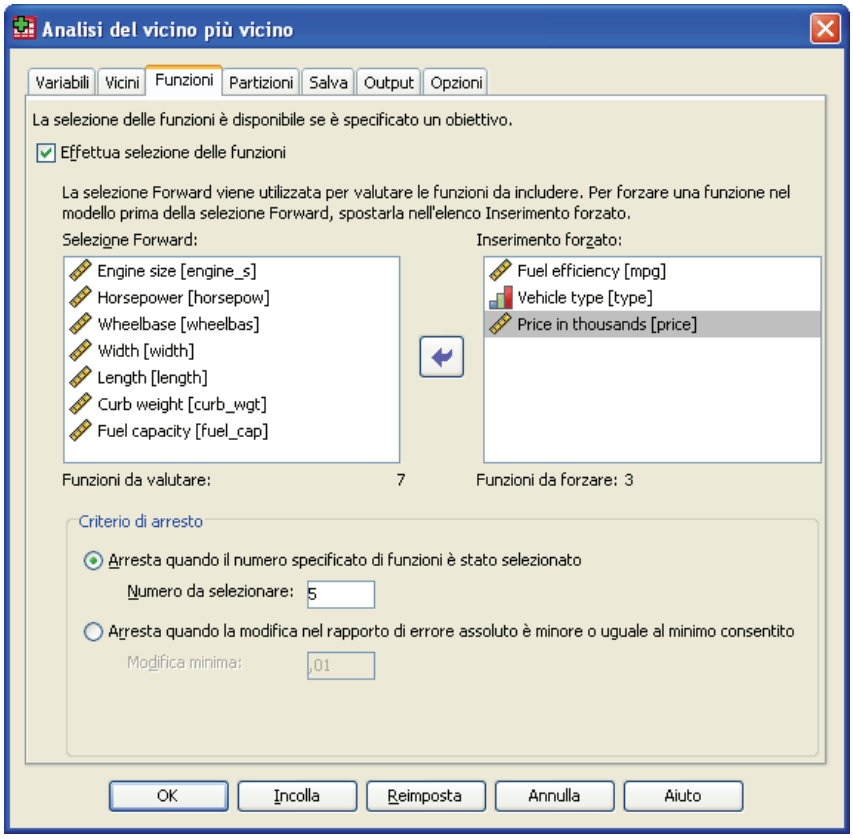
- **Metrica euclidea.** La distanza tra due casi, x e y, è pari alla radice quadrata della somma, in tutte le dimensioni, dei quadrati delle differenze tra i valori di tali casi.
- **Metrica city-block.** La distanza tra due casi è pari alla somma, in tutte le dimensioni, delle differenze assolute tra i valori di tali casi. È denominata anche "distanza di Manhattan".

Se si desidera, qualora nella scheda Variabili sia specificato un obiettivo, è possibile scegliere di ponderare le funzioni in base alla loro importanza normalizzata durante il calcolo delle distanze. L'importanza della funzione di un predittore si calcola dividendo il tasso di errore o l'errore della somma dei quadrati relativo al modello senza il predittore per il tasso di errore o l'errore della somma dei quadrati relativo al modello completo. L'importanza normalizzata si calcola riponderando i valori di importanza della funzione in modo che la somma sia pari a 1.

Previsioni per obiettivo scala. Se nella scheda Variabili è specificato un obiettivo scala, questo indica se il valore previsto è calcolato in base alla media o alla mediana dei vicini più vicini.

Funzioni

Figura 32-4
Finestra di dialogo Analisi del vicino più vicino, scheda Funzioni



La scheda Funzioni consente di richiedere e specificare opzioni per la selezione delle funzioni quando nella scheda Variabili è specificato un obiettivo. Per impostazione predefinita, per la selezione delle funzioni vengono prese in considerazione tutte le funzioni, ma è possibile selezionare un sottoinsieme di funzioni da forzare nel modello.

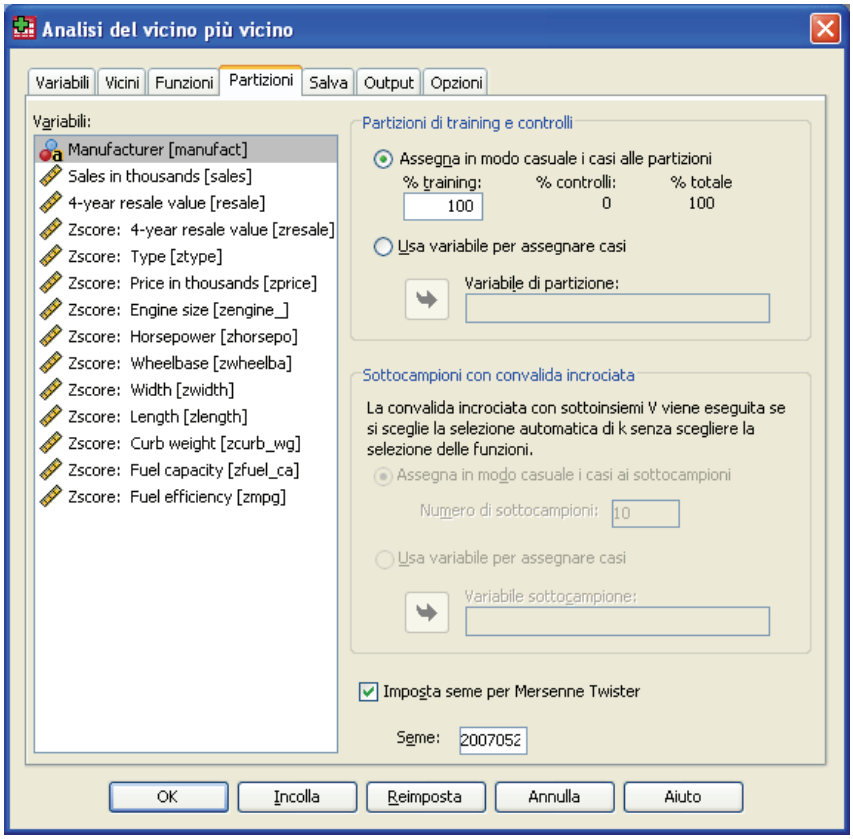
Criterio di arresto. Di volta in volta, viene presa in considerazione, per essere inclusa nell'insieme dei modelli, la funzione il cui inserimento nel modello determina l'errore minore (calcolato come tasso di errore per gli obiettivi categoriali e come errore della somma dei quadrati per gli obiettivi scala). La selezione Forward prosegue fino al raggiungimento della condizione specificata.

- **Numero di funzioni specificato.** L'algoritmo inserisce un numero fisso di funzioni oltre a quelle forzate nel modello. Specificare un intero positivo. La riduzione dei valori del numero da selezionare dà origine a un modello più parsimonioso, con il rischio di perdere funzioni importanti. L'aumento dei valori del numero da selezionare consente di acquisire tutte le funzioni importanti, con il rischio però di aggiungere funzioni che finiscono per moltiplicare l'errore del modello.
- **Variazione minima nel rapporto di errore assoluto.** L'algoritmo si arresta quando la variazione del rapporto di errore assoluto indica che il modello non può essere migliorato ulteriormente aggiungendo altre funzioni. Specificare un numero positivo. La riduzione dei valori della

variazione minima favorisce l'inclusione di un maggior numero di funzioni, con il rischio di inserire funzioni che non aggiungono valore al modello. L'aumento del valore della variazione minima, invece, tende a impedire l'inserimento di altre funzioni, con il rischio di perderne alcune importanti per il modello. Il valore "ottimale" della variazione minima dipende dai dati e dall'applicazione a disposizione. Per assistenza nella valutazione delle funzioni più importanti, vedere il registro degli errori relativi alla selezione delle funzioni nell'output. Per ulteriori informazioni, vedere Registro degli errori relativi alla selezione delle funzioni a pag. 407.

Partizioni

Figura 32-5
Finestra di dialogo Analisi del vicino più vicino, scheda Partizioni



La scheda Partizioni consente di suddividere l'insieme di dati in sottoinsiemi di training e di controllo e, se possibile, di assegnare casi a sottocampioni con convalida incrociata.

Partizioni di training e di controllo. Questo gruppo specifica il metodo di partizionamento dell'insieme di dati attivo in campioni di training e di controllo. Il **campione di training** include i record di dati utilizzati per formare il modello di vicino più vicino; una percentuale di casi nell'insieme di dati deve essere assegnata al campione di training per ottenere un modello. Il **campione di controllo** è un insieme indipendente di record di dati utilizzato per valutare il

modello finale; l'errore per il campione di controllo fornisce una stima "attendibile" della capacità predittiva del modello poiché i casi di controllo non sono stati utilizzati per generare il modello.

- **Assegna in modo casuale i casi alle partizioni.** Specificare la percentuale di casi da assegnare al campione di training. Il resto viene assegnato al campione di controllo.
- **Usa variabile per assegnare casi.** Specificare una variabile numerica che assegni ogni caso nell'insieme di dati attivo al campione di training o di controllo. I casi con valore positivo nella variabile vengono assegnati al campione di training, quelli con valore pari a 0 o negativo al campione di controllo. I casi con un valore di sistema mancante vengono esclusi dall'analisi. I valori mancanti definiti dall'utente per la variabile di partizione sono sempre considerati validi.

Sottocampioni con convalida incrociata. La convalida incrociata con sottocampioni *V* viene utilizzata per determinare il "miglior" numero di vicini. Per motivi legati alle prestazioni, la convalida incrociata non è disponibile se si utilizza la selezione delle funzioni.

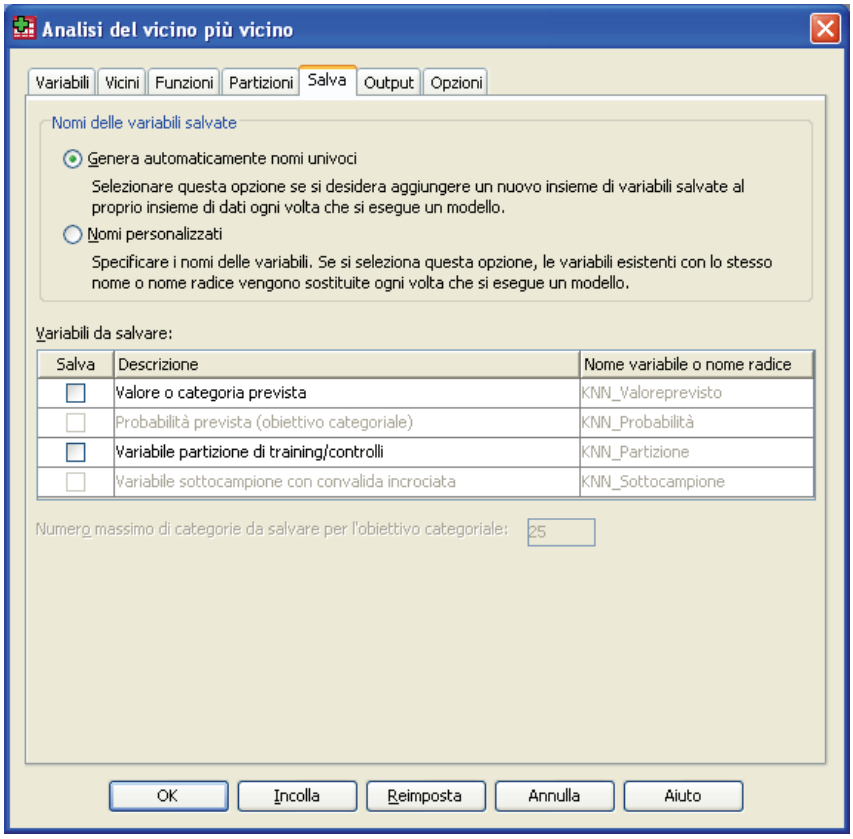
La convalida incrociata suddivide il campione in una serie di sottocampioni. I modelli di vicino più vicino vengono quindi generati escludendo di volta in volta i dati da ciascun sottocampione. Il primo modello si basa su tutti i casi eccetto quelli contenuti nel primo sottocampione, il secondo modello si basa su tutti i casi eccetto quelli contenuti nel secondo sottocampione e così via. Il rischio di errore per ciascun modello viene stimato applicando il modello al sottocampione escluso al momento della generazione del modello stesso. Il "miglior" numero di vicini più vicini è quello che genera l'errore più basso in tutti i sottocampioni.

- **Assegna in modo casuale i casi ai sottocampioni.** Specificare il numero di sottocampioni da utilizzare per la convalida incrociata. I casi vengono assegnati in modo casuale ai sottocampioni, numerati da 1 a *V*, il numero dei sottocampioni.
- **Usa variabile per assegnare casi.** Specificare una variabile numerica che assegni ogni caso nell'insieme di dati attivo a un sottocampione. La variabile deve essere un valore numerico compreso tra 1 e *V*. Se all'interno di tale intervallo mancano valori, e in corrispondenza delle distinzioni in caso di file distinti, si verificherà un errore. Per ulteriori informazioni, vedere Dividi in Capitolo 9 a pag. 192.

Imposta seme per Mersenne Twister. Impostando un seme è possibile replicare le analisi. Questo comando è sostanzialmente un collegamento rapido che consente di impostare Mersenne Twister come il generatore attivo e di specificare un punto di partenza fisso nella finestra di dialogo Generatori di numeri casuali. Per ulteriori informazioni, vedere Generatori di numeri casuali in Capitolo 8 a pag. 139.

Salva

Figura 32-6
Finestra di dialogo Analisi del vicino più vicino, scheda Salva



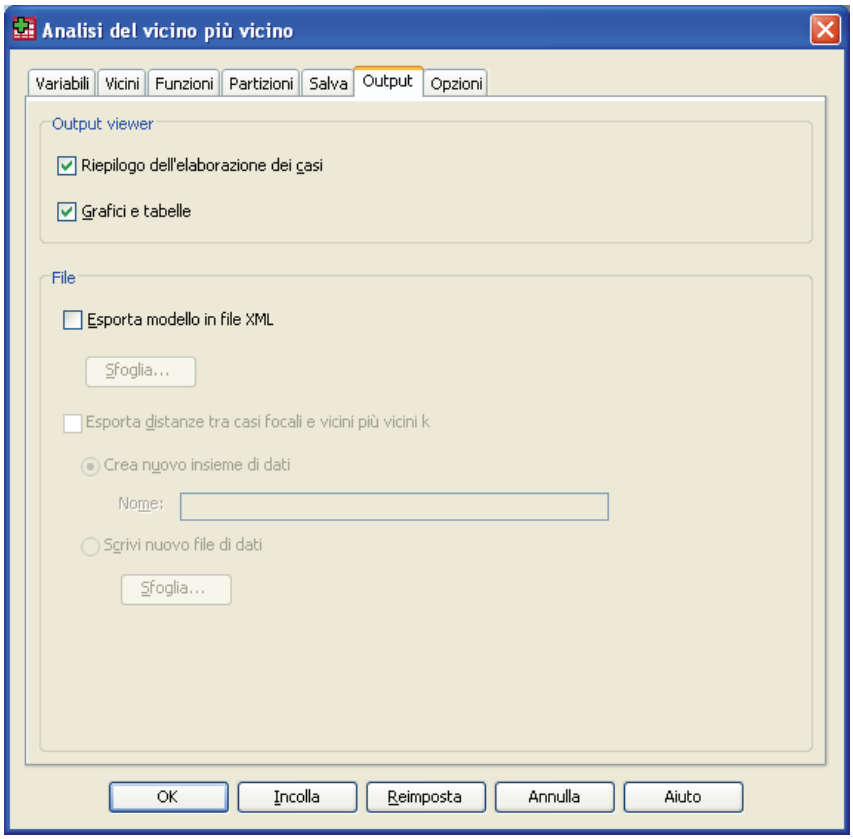
Nomi delle variabili salvate. La generazione automatica del nome assicura il mantenimento di tutto il lavoro. I nomi personalizzati consentono di eliminare/sostituire i risultati di precedenti esecuzioni senza dover prima eliminare le variabili salvate nell'Editor dei dati.

Variabili da salvare

- **Valore o categoria prevista.** Viene salvato il valore previsto per un obiettivo scala o la categoria prevista per un obiettivo categoriale.
- **Probabilità prevista.** Vengono salvate le probabilità previste per un obiettivo categoriale. Una variabile separata viene salvata per ognuna delle prime *n* categorie, dove *n* viene specificato nel comando Numero massimo di categorie da salvare per l'obiettivo categoriale.
- **Variabile partizione di training/controlli.** Se i casi vengono assegnati in modo casuale ai campioni training e di controllo nella scheda Partizioni, viene salvato il valore della partizione (di training o di controllo) a cui il caso è stato assegnato.
- **Variabile sottocampione con convalida incrociata.** Se nella scheda Partizioni ai sottocampioni con convalida incrociata vengono assegnati casi in modo casuale, viene salvato il valore del sottocampione a cui è stato assegnato il caso.

Output

Figura 32-7
Finestra di dialogo Analisi del vicino più vicino, scheda Output



Output Viewer

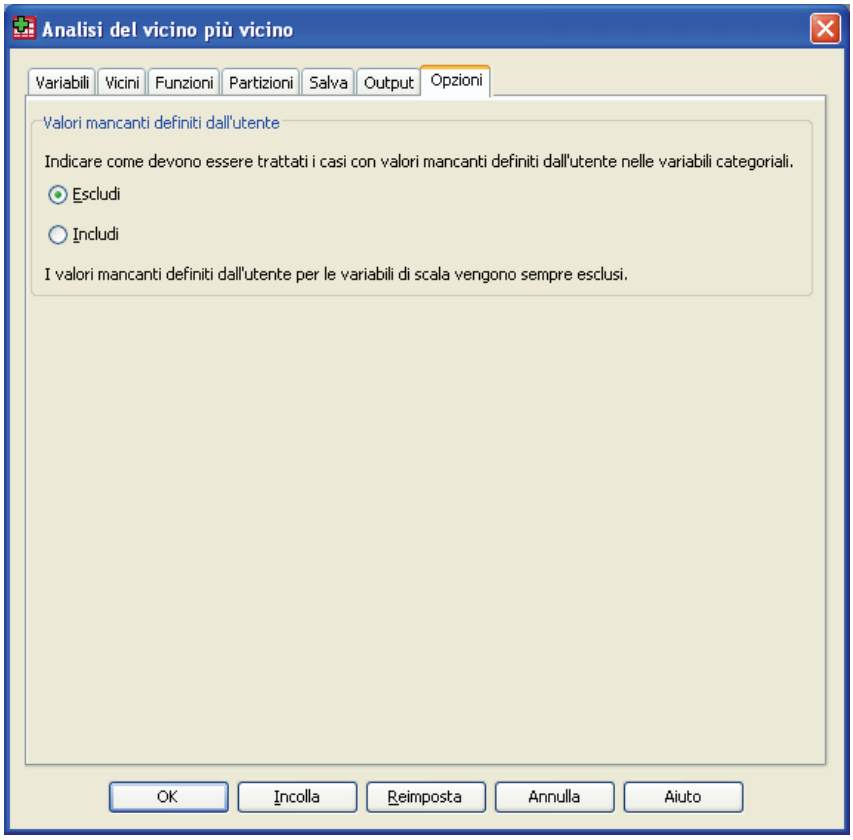
- **Riepilogo dell'elaborazione dei casi.** Visualizza la tabella di riepilogo di elaborazione dei casi, che riepiloga il numero di casi inclusi ed esclusi dall'analisi, in totale e per campioni di training e di controllo.
- **Grafici e tabelle.** Visualizza l'output relativo al modello, tra cui tabelle e grafici. Le tabelle nella vista del modello comprendono i vicini più vicini *k* e le distanze per i casi focali, la classificazione delle variabili di risposta categoriali e un riepilogo degli errori. L'output grafico nella vista del modello include un registro degli errori relativi alla selezione, il grafico dell'importanza delle funzioni, quello dello spazio di funzioni e degli equivalenti e la mappa dei quadranti. Per ulteriori informazioni, vedere Vista del modello a pag. 401.

File

- **Esporta modello in XML.** SmartScore e SPSS Statistics Server (un prodotto distinto) possono utilizzare questo file del modello per applicare le informazioni del modello ad altri file di dati a scopo di assegnazione dei punteggi. Questa opzione non è disponibile se sono stati definiti file suddivisi.
- **Esporta distanze tra casi focali e vicini più vicini k.** Per ogni caso focale viene creata una variabile distinta per ciascuno dei vicini più vicini *k* (dal campione di training) del caso focale stesso e delle corrispondenti distanze più vicine *k*.

Opzioni

Figura 32-8
Finestra di dialogo Analisi del vicino più vicino, scheda Opzioni

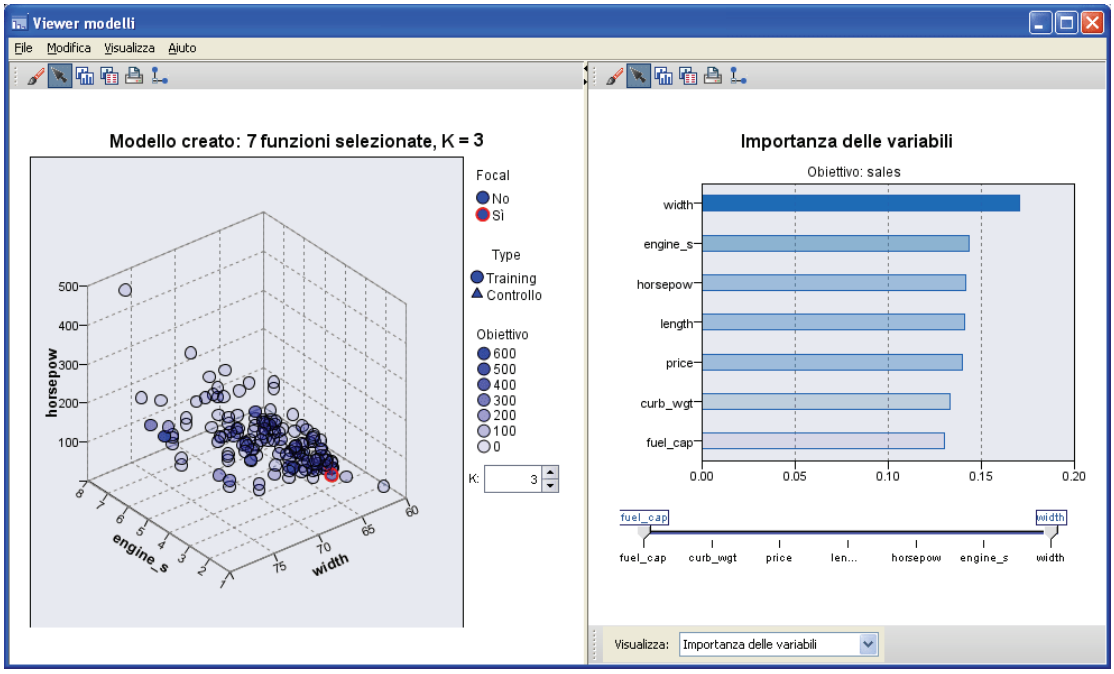


Valori mancanti definiti dall'utente. Le variabili categoriali devono contenere valori validi per un caso per essere incluse nell'analisi. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente devono essere considerati validi nell'ambito delle variabili categoriali.

I valori mancanti di sistema e i valori mancanti relativi alle variabili di scala vengono sempre considerati non validi.

Vista del modello

Figura 32-9
Vista del modello nell'analisi del vicino più vicino



Selezionando Grafici e tabelle nella scheda Output, nel Viewer viene creato un oggetto Modello vicino più vicino. Attivando l'oggetto con un doppio clic, si accede a una vista interattiva del modello con una finestra a due riquadri:

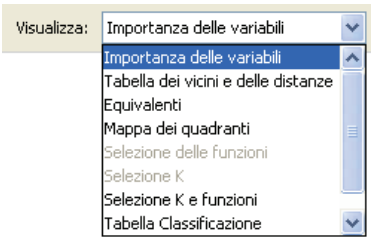
- Nel primo è presente una panoramica del modello denominata "vista principale".
- Nel secondo, invece, possono essere visualizzate due tipologie di vista:

  La vista ausiliaria mostra ulteriori informazioni sul modello, pur non concentrandosi su quest'ultimo.

  La vista collegata mostra invece i dettagli relativi a una funzione del modello quando l'utente esegue il drill-down di parte della vista principale.

Per impostazione predefinita, nel primo riquadro viene visualizzato lo spazio di funzioni e nel secondo il grafico dell'importanza delle variabili. Se quest'ultimo grafico non è disponibile (se, cioè, nella scheda Funzioni non è stato selezionato Pesa funzioni in base all'importanza), viene visualizzata la prima vista presente nell'elenco a discesa Vista.

Figura 32-10
Finestra di dialogo Analisi del vicino più vicino, elenco a discesa Vista del modello

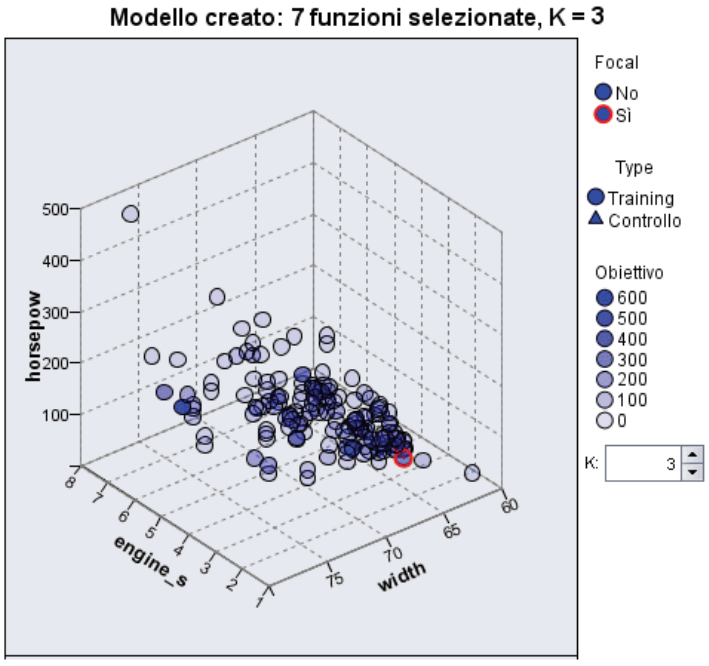


Quando per una vista non sono disponibili informazioni, la voce corrispondente nell'elenco a discesa Vista viene disattivata.

Per informazioni generali sugli oggetti Modello, vedere Modelli a pag. 265

Spazio di funzioni

Figura 32-11
Spazio di funzioni



Il grafico dello spazio di funzioni è un grafico interattivo relativo allo spazio delle funzioni (o al sottospazio, se sono presenti più di tre funzioni). Ogni asse rappresenta una funzione nel modello e la posizione dei punti nel grafico indica i valori di tali funzioni per i casi nelle partizioni di training e di controllo.

Chiavi. Oltre a rappresentare i valori delle funzioni, i punti forniscono altre informazioni.

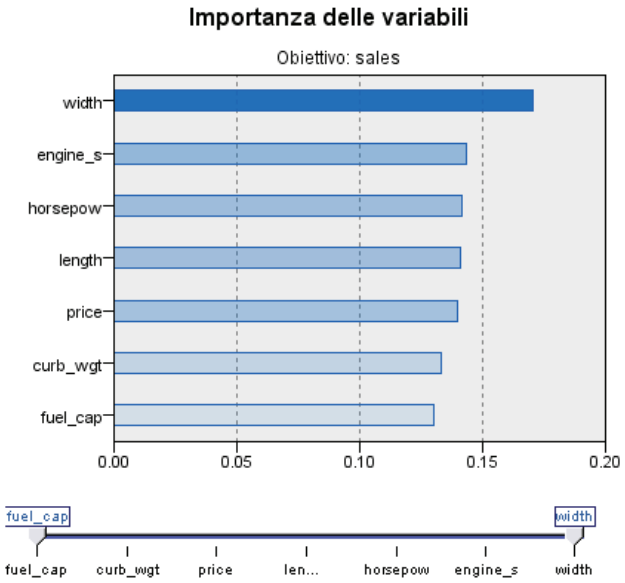
- La forma indica la partizione (Training o Controllo) di cui fa parte un punto. Se non vengono specificate partizioni di controllo, questo simbolo non viene visualizzato.
- Il colore/l'ombreggiatura di un punto indica il valore dell'obiettivo del caso (i diversi valori di colore corrispondono alle categorie di un obiettivo categoriale, mentre le ombreggiature indicano l'intervallo di valori di un obiettivo continuo). Il valore indicato per la partizione di training è quello osservato, mentre per la partizione di controllo è indicato quello previsto. Se non viene specificato alcun obiettivo, questo simbolo non viene visualizzato.
- Contorni più marcati indicano che un caso è focale. I casi focali vengono visualizzati collegati ai relativi vicini più vicini *k*.

Comandi e interattività. Nel grafico è disponibile una serie di comandi per esplorare lo spazio di funzioni.

- È possibile scegliere il sottoinsieme di funzioni da visualizzare nel grafico e cambiare le funzioni da rappresentare nelle dimensioni.
- I casi focali non sono altro che punti selezionati nel grafico relativo allo spazio di funzioni. Se è stata specificata una variabile per casi focali, inizialmente verranno selezionati i punti che rappresentano i casi focali. Qualsiasi punto può comunque diventare temporaneamente un caso focale se viene selezionato. Vengono utilizzati i "soliti" comandi per la selezione di punti (facendo clic su un punto, quest'ultimo viene selezionato e vengono deselezionati tutti gli altri; facendo clic su un punto mentre si tiene premuto il tasto CTRL, tale punto viene aggiunto all'insieme dei punti selezionati). Le viste collegate, ad esempio il grafico degli equivalenti, vengono automaticamente aggiornate in base ai casi selezionati nello spazio di funzioni.
- È possibile modificare il numero di vicini più vicini (*k*) da visualizzare per i casi focali.
- Passando il mouse sopra un punto del grafico, viene visualizzata una descrizione con il valore dell'etichetta del caso (o il numero del caso se non sono state definite etichette), oltre ai valori osservati e previsti dell'obiettivo.
- Il pulsante di ripristino consente di reimpostare lo spazio di funzioni allo stato originario.

Importanza della variabile

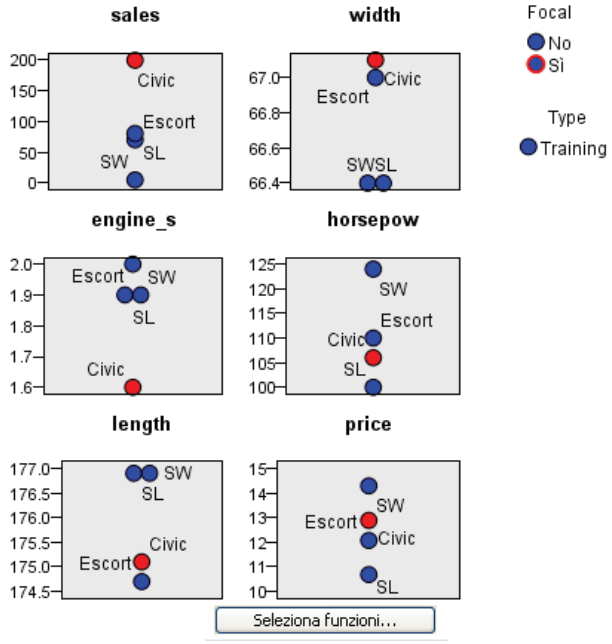
Figura 32-12
Importanza della variabile



Di solito è opportuno concentrare la modellazione sulle variabili più rilevanti, lasciando perdere o ignorando le meno importanti. In questo senso può essere utile il grafico dell'importanza delle variabili, che indica l'importanza relativa di ciascuna variabile nella stima del modello. Dal momento che i valori sono relativi, la somma dei valori di tutte le variabili visualizzate è pari a 1,0. L'importanza delle variabili non ha nulla a che vedere con la precisione del modello. Riguarda unicamente l'importanza di ciascuna variabile per l'elaborazione di una previsione, non il grado di precisione di quest'ultima.

Equivalenti

Figura 32-13
Grafico degli equivalenti



In questo grafico vengono visualizzati i casi focali e i relativi vicini più vicini *k* per ciascuna funzione e per l'obiettivo. È disponibile se nello spazio di funzioni è selezionato un caso focale.

Collegamenti. Il grafico degli equivalenti è collegato allo spazio di funzioni in due modi.

- I casi selezionati (focali) nello spazio di funzioni vengono visualizzati nel grafico degli equivalenti, insieme ai relativi vicini più vicini *k*.
- Il valore di *k* selezionato nello spazio di funzioni viene utilizzato nel grafico degli equivalenti.

Distanze dei vicini più vicini

Figura 32-14
Distanze dei vicini più vicini

Caso focale	Vicini più vicini			Distanze più prossime		
	1	2	3	1	2	3
Civic	SL	Escort	SW	0.053674	0.0599112	0.0643422

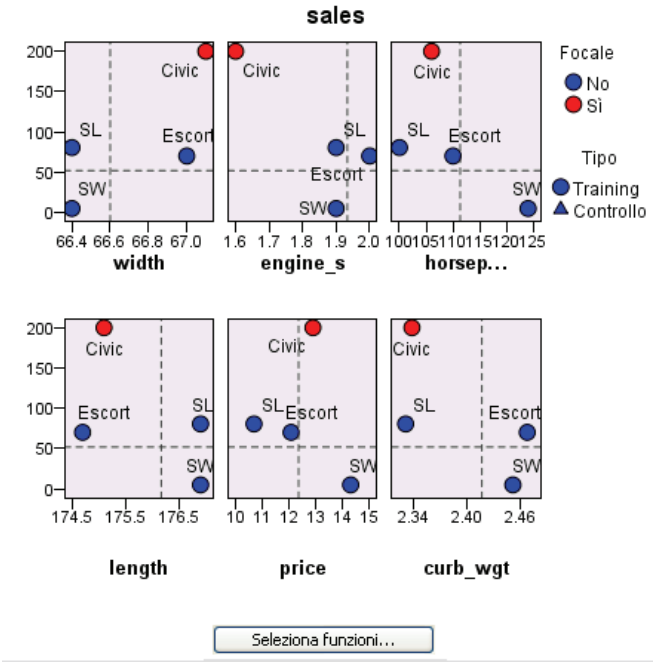
In questa tabella vengono visualizzati i vicini più vicini e le distanze più vicine *k* solo per i casi focali. È disponibile se nella scheda Variabili è specificato un identificatore dei casi focali e mostra soltanto i casi focali identificati da questa variabile.

Ogni riga della:

- Colonna Caso focale contiene il valore della variabile di etichetta relativa al caso focale. Se non sono definite etichette dei casi, la colonna contiene il numero di caso del caso focale.
- *i*ª colonna del gruppo Vicini più vicini contiene il valore della variabile di etichetta dei casi relativa al *i*° vicino più vicino del caso focale. Se non sono definite etichette dei casi, la colonna contiene il numero di caso del *i*° vicino più vicino del caso focale.
- *i*ª colonna del gruppo Distanze più vicine contiene la distanza del *i*° vicino più vicino dal caso focale.

Mappa dei quadranti

Figura 32-15
Mappa dei quadranti

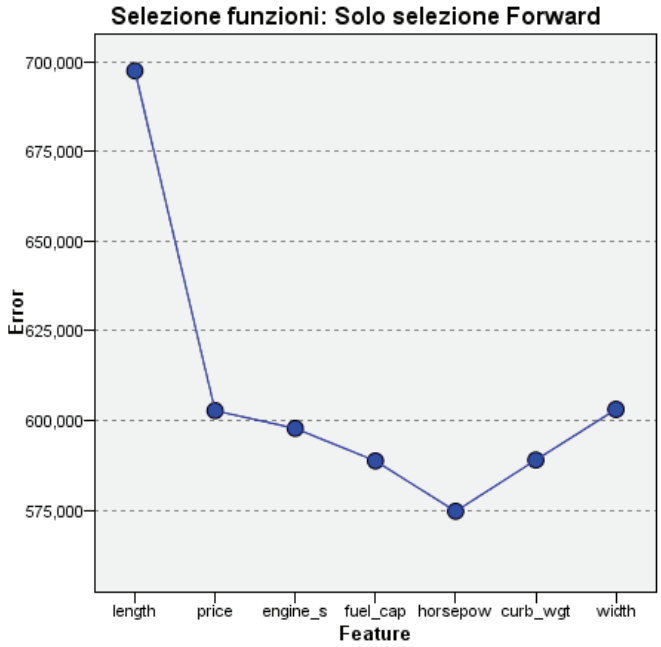


Il grafico mostra i casi focali e i relativi vicini più vicini *k* su un grafico a dispersione (o un grafico a punti a seconda del livello di misurazione dell'obiettivo) con l'obiettivo sull'asse *y* e una funzione di scala sull'asse *x*, il tutto suddiviso in riquadri in base alle funzioni. È disponibile se nello spazio di funzioni è presente un obiettivo ed è selezionato un caso focale.

- Per le variabili continue, nella partizione di training in corrispondenza delle medie delle variabili vengono tracciate linee di riferimento.

Registro degli errori relativi alla selezione delle funzioni

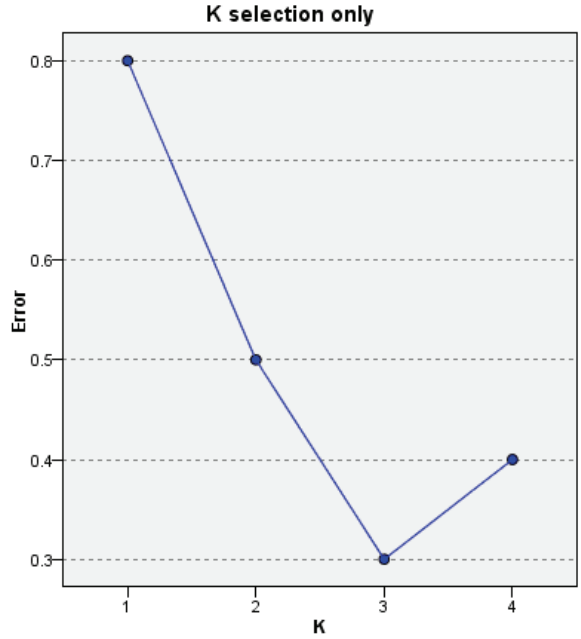
Figura 32-16
Selezione delle funzioni



I punti presenti nel grafico mostrano l'errore (in termini di tasso di errore o di errore della somma dei quadrati a seconda del livello di misurazione dell'obiettivo) sull'asse *y* del modello, con la funzione elencata sull'asse *x* (inoltre, a sinistra sull'asse *x* sono presenti tutte le funzioni). Il grafico è disponibile se è presente un obiettivo ed è attiva la selezione funzioni.

Registro degli errori relativi alla selezione k

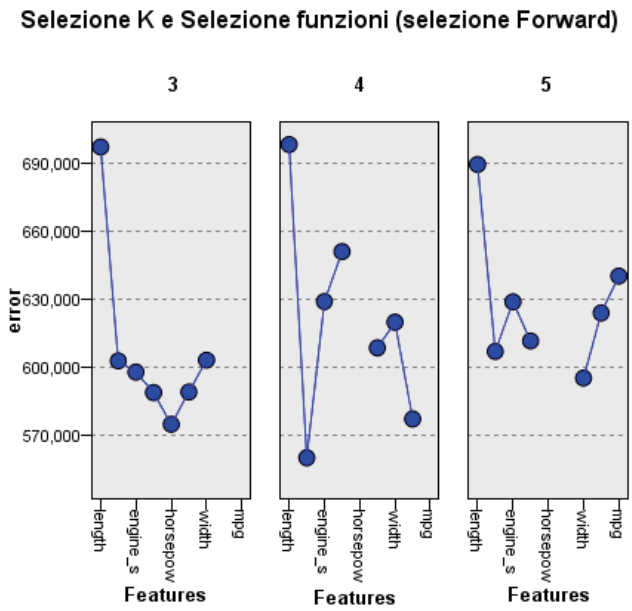
Figura 32-17
Selezione k



I punti presenti nel grafico mostrano l'errore (in termini di tasso di errore o di errore della somma dei quadrati a seconda del livello di misurazione dell'obiettivo) sull'asse *y* del modello, con il numero di vicini più vicini (*k*) sull'asse *x*. Il grafico è disponibile se è presente un obiettivo ed è attiva la selezione *k*.

Registro degli errori relativi alla selezione k e alla selezione delle funzioni

Figura 32-18
Selezione k e selezione delle funzioni



Si tratta di grafici per la selezione delle funzioni (vedere Registro degli errori relativi alla selezione delle funzioni a pag. 407), suddivisi in riquadri in base a *k*. Il grafico è disponibile se è presente un obiettivo e sono attive sia la selezione *k* sia la selezione delle funzioni.

Tabella di classificazione

Figura 32-19
Tabella di classificazione

Partizione		Previsto		
		0	1	Percentuale corretta
Training	0	111	1	99.11%
	1	7	33	82.50%
	Percentuale globale	77.64%	22.37%	94.74%

Nella tabella viene visualizzata la classificazione incrociata dei valori osservati dell'obiettivo rispetto a quelli previsti, suddivisi per partizione. È disponibile se è presente un obiettivo di tipo categoriale.

- La riga (Mancante) della partizione di controllo contiene casi di controllo con valori mancanti sull'obiettivo. Tali casi contribuiscono ai valori di percentuale complessiva, ma non a quelli di percentuale corretta, del campione di controllo.

Riepilogo degli errori

Figura 32-20
Riepilogo degli errori

Partizione	Sum-of-Squares Error
Training	622043

La tabella è disponibile in presenza di una variabile di destinazione. Mostra l'errore associato al modello: la somma dei quadrati per l'obiettivo continuo e il tasso di errore (percentuale complessiva di correttezza del 100%) per un obiettivo categoriale.

Capitolo
33

Analisi discriminante

L'analisi discriminante crea un modello predittivo per l'appartenenza ai gruppi. Il modello è composto da una funzione discriminante (oppure, per più di due gruppi, da un insieme di funzioni discriminanti) basata su combinazioni lineari delle variabili predittore che forniscono la discriminazione ottimale tra i gruppi. Le funzioni vengono generate da un campione di casi di cui è nota l'appartenenza; le funzioni possono in seguito essere applicate ai nuovi casi che hanno misurazioni delle variabili predittore ma la cui appartenenza di gruppo è sconosciuta.

Nota: La variabile di raggruppamento può includere più di due valori. I codici per la variabile di raggruppamento devono tuttavia essere interi, ed è necessario specificare i valori massimo e minimo corrispondenti. I casi con valore non compreso tra i due estremi specificati vengono esclusi dall'analisi.

Esempio. In media, il consumo calorico giornaliero degli abitanti delle zone temperate è maggiore di quello di chi vive ai tropici. Nelle zone temperate, inoltre, si riscontra una maggiore percentuale di persone che vivono in ambiente urbano. Un ricercatore desidera combinare queste informazioni in una funzione per determinare le modalità di discriminazione tra i due gruppi di paesi. Il ricercatore ritiene opportuno prendere in considerazione anche le dimensioni della popolazione e informazioni di carattere economico. L'analisi discriminante consente di valutare i coefficienti della funzione discriminante lineare, analoga alla parte destra di un'equazione di regressione lineare multipla. In altri termini, utilizzando i coefficienti *a*, *b*, *c* e *d* si ottiene la funzione:

D = a * clima + b * urbano + c * popolazione + d * prodotto interno lordo pro capite

Se queste variabili sono utili per la discriminazione tra le due zone climatiche, i valori di *D* per i paesi temperati saranno diversi da quelli relativi ai paesi tropicali. Se è necessario usare un metodo di selezione delle variabili per passi, nella funzione non si dovranno includere tutte e quattro le variabili.

Statistiche. Per ogni variabile: media, deviazione standard, ANOVA univariata. Per ogni analisi: *M* di Box, matrice di correlazione entro gruppi, matrice di covarianza entro gruppi, matrice di covarianza di gruppi separati e matrice di covarianza totale. Per ciascuna funzione discriminante canonica: autovalori, percentuale di varianza, correlazione canonica, lambda di Wilks, chi-quadrato. Per ogni passo: probabilità a priori, coefficienti di funzione di Fisher, coefficienti di funzione standardizzati, lambda di Wilks per ogni funzione canonica.

Dati. La variabile di raggruppamento deve includere un numero limitato di categorie distinte, codificate come interi. Le variabili indipendenti nominali devono essere ricodificate in forma di variabili fittizie o di contrasto.

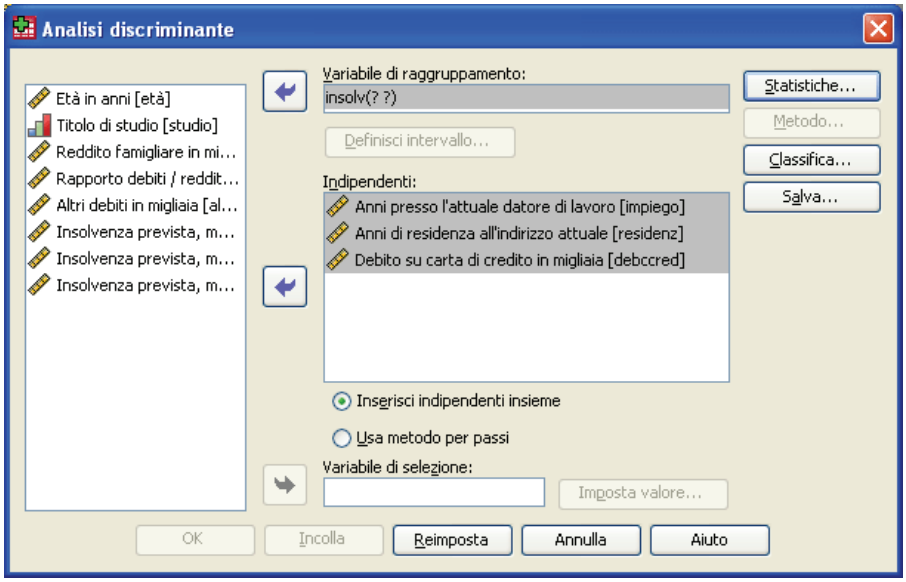
Assunzioni. I casi devono essere indipendenti. Le variabili stimatore devono avere una distribuzione normale multivariata e le matrici di varianza-covarianza entro gruppi devono essere uguali in tutti i gruppi. Si assume che le appartenenze ai gruppi si escludano reciprocamente (ovvero che nessun caso appartenga a più gruppi) e che ciascun caso appartenga a un gruppo. La procedura è più efficace se l'appartenenza ai gruppi è una variabile categoriale effettiva; se l'appartenenza ai gruppi si basa sui valori di una variabile continua (ad esempio, QI massimo e QI minimo), è opportuno utilizzare la regressione lineare per avvalersi delle informazioni più dettagliate disponibili nella variabile continua.

Per ottenere un'analisi discriminante

▶ Dai menu, scegliere:

Analizza
  Classifica
    Discriminante...

Figura 33-1
Finestra di dialogo Analisi discriminante



▶ Selezionare una variabile di raggruppamento con valori interi e fare clic su Definisci intervallo per specificare le categorie desiderate.

▶ Selezionare le variabili indipendenti o le variabili stimatore. (Se la variabile di raggruppamento non include valori interi, utilizzando il comando Ricodifica automatica del menu Trasforma è possibile crearne una che includa tali valori).

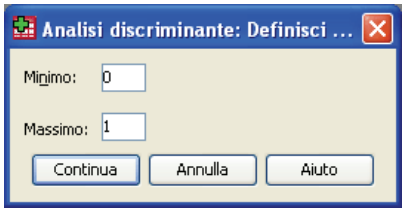
▶ Selezionare il metodo di inserimento delle variabili indipendenti.

- **Inserisci indipendenti insieme.** Inserisce contemporaneamente tutte le variabili indipendenti che soddisfano i criteri di tolleranza.
- **Usa metodo stepwise.** Usa un metodo per passi per controllare l'inserimento e la rimozione delle variabili.

- È inoltre possibile selezionare i casi utilizzando una variabile di selezione.

Analisi discriminante: Definisci intervallo

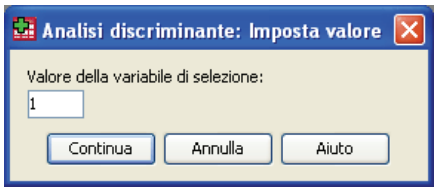
Figura 33-2
Finestra di dialogo Analisi discriminante: Definisci intervallo



Specificare il valore minimo e massimo della variabile di raggruppamento da utilizzare per l'analisi. I casi con valori che non rientrano in tale intervallo non vengono utilizzati nell'analisi discriminante, ma vengono classificati in uno dei gruppi esistenti in base ai risultati dell'analisi stessa. I valori massimo e minimo devono essere interi.

Analisi discriminante: Seleziona casi

Figura 33-3
Finestra di dialogo Analisi discriminante: Imposta valore



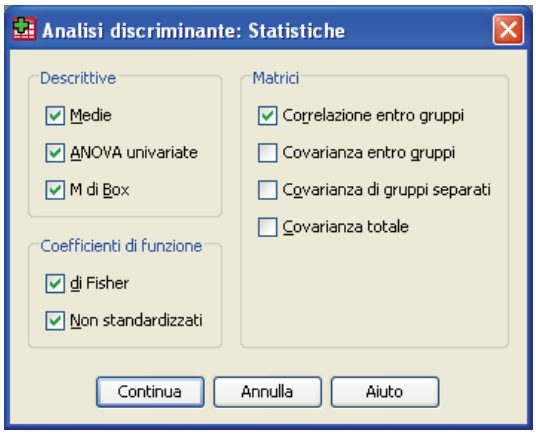
Per selezionare i casi da usare nell'analisi:

- Scegliere la variabile di selezione nella finestra di dialogo Analisi discriminante.
- Fare clic suValore per immettere la variabile di selezione come numero intero.

Le funzioni discriminanti verranno derivate solo in base ai casi che prevedono tale valore per la variabile di selezione. Le statistiche e i risultati della classificazione vengono generati sia per i casi selezionati che per i casi non selezionati. Tale processo consente di classificare i nuovi casi sulla base dei dati preesistenti nonché ripartire i dati in sottoinsiemi di prova con i quali eseguire la convalida del modello generato.

Analisi discriminante: Statistiche

Figura 33-4
Finestra di dialogo Analisi discriminante: Statistiche



Descrittive. Le opzioni disponibili sono medie (incluse le deviazioni standard), ANOVA univariate e test *M* di Box.

- **Medie.** Visualizza le medie e le deviazioni standard globali e di gruppo delle variabili indipendenti.
- **ANOVA univariate.** Effettua l'analisi della varianza univariata per verificare l'uguaglianza delle medie di gruppo per ciascuna variabile indipendente.
- **M di Box.** Un test per l'uguaglianza di matrici di covarianza di gruppo. Per dimensioni di campione sufficientemente elevate, un valore P non significativo vuol dire che non ci sono sufficienti prove che le matrici differiscano. Il test è sensibile a scostamenti dalla normalità multivariata, tende cioè a non considerare uguali le matrici se l'ipotesi di normalità viene violata.

Coefficienti di funzione. Le opzioni disponibili sono i coefficienti di correlazione di Fisher e i coefficienti non standardizzati.

- **Di Fisher (Analisi discriminante).** Visualizza i coefficienti di Fisher della funzione discriminante, che possono essere usati direttamente per la classificazione. Viene riprodotto un insieme separato di coefficienti di funzioni di classificazione per ciascun gruppo. Ogni caso viene assegnato al gruppo in cui ottiene il più alto punteggio discriminante (valore della funzione di classificazione).
- **Non standardizzati.** Visualizza i coefficienti della funzione discriminante non standardizzati.

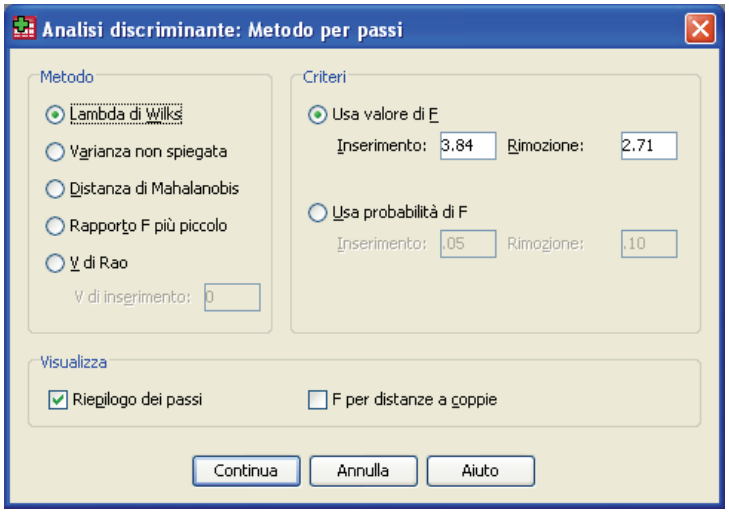
Matrici. Le matrici di coefficienti disponibili per le variabili indipendenti sono: matrice di correlazione entro gruppi, matrice di covarianza entro gruppi, matrice di covarianza gruppi separati e matrice di covarianza totale.

- **Correlazione entro gruppi.** Visualizza la matrice di correlazione entro gruppi ottenuta mediando le matrici di covarianza di tutti i gruppi prima di calcolare le correlazioni.
- **Covarianza entro gruppi.** Visualizza una matrice combinata di covarianza entro i gruppi, mediando le singole matrici di covarianza di tutti i gruppi. Questa matrice può essere diversa dalla matrice di covarianza globale.

- **Covarianza di gruppi separati.** Visualizza matrici di covarianza separate per ciascun gruppo.
- **Covarianza totale.** Visualizza la matrice di covarianza globale, calcolata su tutti i casi, ovvero ignorando la suddivisione in gruppi.

Analisi discriminante: Metodo Stepwise

Figura 33-5
Finestra di dialogo Analisi discriminante: Metodo Stepwise



Metodo. Selezionare la statistica da utilizzare per l'inserimento o la rimozione di nuove variabili. Le alternative disponibili sono lambda di Wilks, varianza non spiegata, distanza di Mahalanobis, minimo rapporto *F* e *V* di Rao. Con il *V* di Rao è possibile specificare l'aumento minimo in *V* per la variabile da inserire.

- **Lambda di Wilks.** Un metodo di selezione delle variabili nell'analisi discriminante per passi che sceglie le variabili da inserire nell'equazione in base a quanto esse contribuiscono a minimizzare il Lambda di Wilks. Ad ogni passo viene inserita la variabile che minimizza il valore globale del Lambda di Wilks'.
- **Varianza non spiegata.** Ad ogni passo viene inserita la variabile che riduce al minimo la somma della variazione spiegata fra gruppi.
- **Distanza di Mahalanobis.** Una misura della distanza di un caso dalla media di tutti i casi per le variabili indipendenti. Un'elevata distanza di Mahalanobis indica che un caso include valori estremi per una o più variabili indipendenti.
- **Rapporto F più piccolo.** Un metodo di selezione delle variabili nelle analisi per passi basato sulla massimizzazione di un rapporto F valutato tramite la distanza di Mahalanobis tra gruppi.
- **V di Rao.** Una misura delle differenze tra medie di gruppo. Detta anche traccia di Lawley-Hotelling. Ad ogni passo viene inserita la variabile che massimizza l'aumento della V di Rao. Specificare l'incremento minimo che una variabile deve apportare per essere inserita nel modello.

Criteri. Le alternative disponibili sono Usa valore di F e Usa probabilità di F. Immettere valori per aggiungere o rimuovere variabili.

- **Usa valore di F.** La variabile viene inserita nel modello se il relativo valore F è maggiore di quello di inserimento. La variabile viene altresì rimossa se il relativo valore F è minore di quello di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi e Inserimento deve essere maggiore di Rimozione Abbassando il valore di inserimento e/o alzando quello di rimozione si allentano i vincoli di inclusione delle variabili nel modello.
- **Usa probabilità di F.** La variabile viene inserita nel modello se il livello di significatività del relativo valore di F è minore di quello di inserimento. La variabile viene altresì rimossa se il livello di significatività è maggiore di quello di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi e Inserimento deve essere minore di Rimozione. Alzando il valore di inserimento e/o abbassando quello di rimozione si allentano i vincoli di inclusione delle variabili nel modello.

Visualizzazione. Riepilogo dei passi visualizza le statistiche per tutte le variabili dopo ogni passo. F per distanze a coppie visualizza una matrice di valori *F* pairwise per ciascuna coppia di gruppi.

Analisi discriminante: Classificazione

Figura 33-6
Finestra di dialogo Analisi discriminante: Classificazione

Probabilità a priori. Questa opzione determina se i coefficienti di classificazione vengono corretti per una conoscenza a priori del gruppo di appartenenza.

- **Tutti i gruppi uguali.** Si presuppongono probabilità a priori uguali per tutti i gruppi, senza effetti sui coefficienti.
- **Calcola dalle dimensioni dei gruppi.** Le dimensioni del gruppo osservate determinano le probabilità a priori del gruppo di appartenenza. Ad esempio, se il 50% delle osservazioni incluse nell'analisi rientrano nel primo gruppo, il 25% nel secondo e il 25% nel terzo, i coefficienti di classificazione vengono corretti in modo da aumentare la probabilità di appartenenza nel primo gruppo rispetto agli altri due.

Visualizzazione. Le opzioni di visualizzazione disponibili sono: risultati per casi, tabella riassuntiva e classificazione autoesclusiva.

- **Risultati per casi.** Visualizza per ciascun caso i codici del gruppo effettivo, del gruppo previsto, della probabilità a posteriori e del punteggio discriminante.
- **Tabella riassuntiva.** Il numero di casi assegnati in modo corretto e non corretto a ciascuno dei gruppi in base all'analisi discriminante. A volte detta "Matrice confusione".
- **Classificazione autoesclusiva.** Ogni caso viene classificato usando le funzioni costruite su tutti i casi meno se stesso. Nota anche come classificazione leave one out o metodo U.

Sostituisci valori mancanti con la media. Selezionare questa opzione per sostituire un valore mancante con la media di una variabile indipendente, solo durante la fase di classificazione.

Usa matrice di covarianza. È possibile decidere di classificare i casi utilizzando una matrice di covarianza entro i gruppi o una matrice di covarianza per gruppi separati.

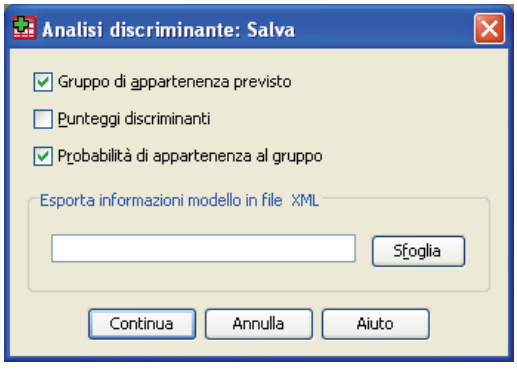
- **Entro i gruppi.** Per classificare i casi viene utilizzata la matrice globale di covarianza entro i gruppi.
- **Gruppi separati.** Per classificare i casi vengono utilizzate le matrici di covarianza dei singoli gruppi. Dal momento che la classificazione è basata sulla funzione discriminante e non sui valori originali, questa opzione non è sempre equivalente alla discriminazione quadratica.

Grafici. Le opzioni disponibili per i grafici sono: gruppi accorpati, gruppi separati e mappa territoriale.

- **Gruppi accorpati.** Produce un unico grafico a dispersione dei valori delle prime due funzioni discriminanti. Se c'è una sola funzione, viene prodotto un istogramma.
- **Gruppi separati.** Produce un grafico a dispersione dei valori delle prime due funzioni discriminanti per ciascun gruppo. Se c'è una sola funzione, verranno prodotti istogrammi.
- **Mappa territoriale.** Un grafico dei confini usati per classificare i casi in gruppi in base ai valori di una funzione. I numeri corrispondono ai gruppi nei quali vengono classificati i casi. La media per ciascun gruppo è indicata da un asterisco all'interno dei suoi confini. La mappa non viene visualizzata se c'è una sola funzione discriminante.

Analisi discriminante: Salva

Figura 33-7
Finestra di dialogo Analisi discriminante: Salva



È possibile aggiungere nuove variabili al file di dati attivo. Le opzioni disponibili sono: gruppo di appartenenza previsto (una sola variabile), punteggi discriminanti (una variabile per ciascuna funzione discriminante nella soluzione) e probabilità di appartenenza al gruppo in base ai punteggi discriminanti (una variabile per ciascun gruppo).

È possibile esportare le informazioni sul modello nel file specificato in formato XML (PMML). SmartScore e SPSS Statistics Server (un prodotto separato) possono utilizzare questo file del modello per applicare le informazioni del modello ad altri file di dati a scopo di assegnazione dei punteggi.

Funzioni aggiuntive del comando DISCRIMINANT

Il linguaggio della sintassi dei comandi consente inoltre di:

- Eseguire più analisi discriminanti (con un comando) e controllare l'ordine in cui le variabili vengono inserite (con il sottocomando `ANALYSIS`).
- Specificare le probabilità a priori per la classificazione (con il sottocomando `PRIORS`).
- Visualizzare le matrici dei modelli e delle strutture ruotate (con il sottocomando `ROTATE`).
- Limitare il numero di funzioni discriminanti estratte (con il sottocomando `FUNCTIONS`).
- Limitare la classificazione ai casi selezionati (o non selezionati) per l'analisi (con il sottocomando `SELECT`).
- Leggere e analizzare la matrice di correlazione (con il sottocomando `MATRIX`).
- Scrivere una matrice di correlazione da analizzare in seguito (con il sottocomando `MATRIX`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
34

Analisi fattoriale

L'analisi fattoriale si propone di identificare le variabili sottostanti, o **fattori**, che illustrano il modello per le correlazioni all'interno di un insieme di variabili osservate. L'analisi fattoriale viene in genere utilizzata per la riduzione dei dati in quanto consente di identificare un numero ridotto di valori che spiegano la maggior parte dei valori di varianza osservati in numerose variabili manifeste. L'analisi fattoriale può inoltre essere utilizzata per generare ipotesi relative a meccanismi causali oppure per esaminare le variabili per le analisi successive (ad esempio per identificare la collinearità prima di eseguire un'analisi di regressione lineare).

La procedura di analisi fattoriale permette un elevato grado di flessibilità:

- Sono disponibili sette metodi di estrazione fattoriale.
- Sono disponibili cinque metodi di rotazione, tra cui oblimin diretto e promax per le rotazioni non ortogonali.
- Sono disponibili tre metodi per il calcolo dei punteggi, che possono essere salvati come variabili per le analisi successive.

Esempio. Quali sono gli atteggiamenti sottostanti che inducono le persone a rispondere a questionari politici in un determinato modo? Dall'esame delle correlazioni esistenti tra le voci del questionario risulta una significativa sovrapposizione tra diversi sottogruppi di voci. Ad esempio, le domande relative alle tasse tendono ad essere correlate fra loro, così come le domande relative alle questioni militari e così via. Grazie all'analisi fattoriale è possibile identificare il numero di fattori sottostanti e in molti casi determinare cosa rappresentano concettualmente tali fattori. È inoltre possibile calcolare i punteggi fattoriali per ciascun rispondente, un elemento che è possibile utilizzare in analisi successive. Ad esempio, è possibile creare un modello di regressione logistico che consenta di prevedere il comportamento di voto in base ai punteggi fattoriali.

Statistiche. Per ogni variabile: numero di casi validi, media e deviazione standard. Per ciascuna analisi fattoriale: matrice di correlazione delle variabili, inclusi i livelli di significatività, determinante, inversa; matrice di correlazione riprodotta, inclusa anti-immagine; soluzione iniziale (comunalità, autovalori e percentuale di varianza spiegata); misura di adeguatezza campionaria di Kaiser-Meyer-Olkin e test di sfericità di Bartlett; soluzione non ruotata, inclusi pesi fattoriali, comunalità e autovalori; soluzione ruotata, incluse la matrice ruotata dei modelli e la matrice di trasformazione. Per le rotazioni oblique: matrice ruotata dei modelli e delle strutture; matrice dei coefficienti di punteggio fattoriale e matrice di covarianza fattoriale. Grafici: grafico decrescente degli autovalori e grafico dei pesi fattoriali dei primi due o tre fattori.

Dati. Le variabili devono essere quantitative al livello di misura per **intervallo** o per **rapporto**. I dati categoriali (ad esempio religione o paese d'origine) non sono idonei per l'analisi fattoriale. I dati per cui è possibile calcolare i coefficienti di correlazione di Pearson sono idonei per l'analisi fattoriale.

Assunzioni. I dati devono avere una distribuzione normale bivariata per ciascuna coppia di variabili e le osservazioni devono essere indipendenti. Il modello di analisi fattoriale specifica che le variabili vengono determinate da fattori comuni (i fattori stimati dal modello) e fattori univoci (che non risultano sovrapposti tra le variabili osservate); le stime calcolate si basano sull'ipotesi che tutti i fattori univoci siano correlati reciprocamente e con i fattori comuni.

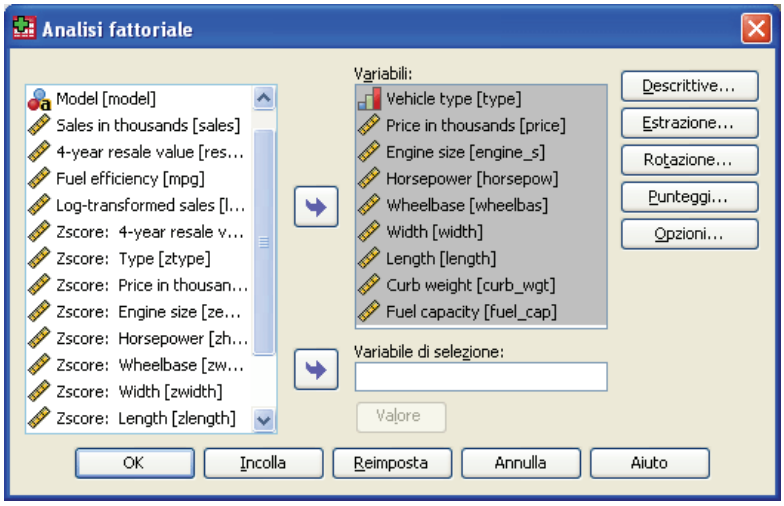
Per ottenere un'analisi fattoriale

- Dai menu, scegliere:

  Analizza
  Riduzioni dimensione
  Fattoriale...

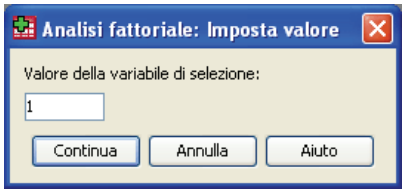
- Selezionare le variabili per l'analisi fattoriale.

Figura 34-1
Finestra di dialogo Analisi fattoriale



Analisi fattoriale: Seleziona casi

Figura 34-2
Finestra di dialogo Analisi fattoriale: Imposta valore



Per selezionare i casi da usare nell'analisi:

- Selezionare una variabile di selezione.
- Fare clic suValore per immettere la variabile di selezione come numero intero.

  Nell'analisi fattoriale verranno utilizzati solo i casi con tale valore per la variabile di selezione.

Analisi fattoriale: Descrittive

Figura 34-3
Finestra di dialogo Analisi fattoriale: Descrittive

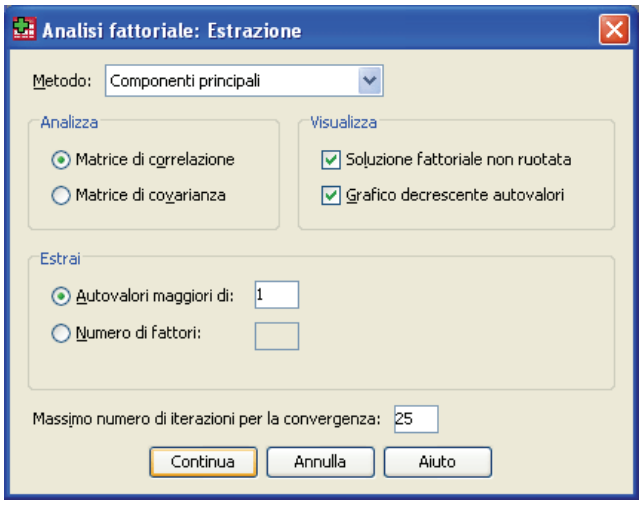
Statistiche. Le statistiche descrittive univariate includono la media, la deviazione standard e il numero di casi validi per ciascuna variabile. Nella soluzione iniziale vengono visualizzate le comunalità iniziali, gli autovalori e la percentuale della varianza spiegata.

Matrice di correlazione. Le opzioni disponibili sono: coefficienti, livelli di significatività, determinante, inversa, riprodotta, anti-immagine, test KMO e test di sfericità di Bartlett.

- **Test KMO e test di sfericità di Bartlett.** La misura di adeguatezza campionaria KMO (Keiser Meyer Olkin) verifica se le correlazioni parziali tra le variabili sono piccole. Il test di sfericità di Bartlett verifica se la matrice di correlazione è una matrice identità, cosa che indicherebbe l'inadeguatezza del modello fattoriale.
- **Riprodotta.** La matrice di correlazione stimata a partire dalla soluzione del fattore. Vengono visualizzati anche i residui (differenze tra correlazioni stimate e osservate).
- **Anti-immagine.** La matrice di correlazione anti-immagine contiene i coefficienti di correlazione parziale cambiati di segno e la matrice di covarianza anti-immagine contiene le covarianze parziali cambiate di segno. In un buon modello fattoriale, la maggior parte degli elementi fuori dalla diagonale dovrebbe avere valori bassi. La misura di adeguatezza campionaria di una variabile è visualizzata sulla diagonale della matrice di correlazione anti-immagine.

Analisi fattoriale: Estrazione

Figura 34-4
Finestra di dialogo Analisi fattoriale: Estrazione



Metodo. Consente di specificare il metodo di estrazione dei fattori. I metodi disponibili sono: componenti principali, minimi quadrati non ponderati, minimi quadrati generalizzati, massima verosimiglianza, fattorizzazione dell'asse principale, fattorizzazione alfa e fattorizzazione immagine.

- **Componenti principali (Analisi fattoriale).** Metodo usato per formare combinazioni lineari non correlate delle variabili osservate. La prima componente spiega la parte più alta di variabilità. Le componenti successive spiegano porzioni di variabilità decrescenti e sono tutte non correlate fra loro. L'analisi delle componenti principali viene usata per ottenere la soluzione fattoriale iniziale. Può essere usata quando una matrice di correlazione è singolare.
- **Metodo dei minimi quadrati non ponderati.** Un metodo per l'estrazione dei fattori che minimizza la somma dei quadrati delle differenze tra la matrice di correlazione osservata e quella riprodotta, ignorando le diagonali.
- **Minimi quadrati generalizzati (Analisi fattoriale).** Un metodo di estrazione dei fattori che minimizza la somma dei quadrati delle differenze tra la matrice di correlazione osservata e la matrice di correlazione riprodotta. Le correlazioni sono ponderate tramite l'inverso della loro unicità, in modo da dare meno peso alle variabili con elevata unicità rispetto a quelle con unicità inferiore.
- **Massima verosimiglianza (Analisi fattoriale).** Un metodo per l'estrazione dei fattori che produce le stime dei parametri che più verosimilmente hanno prodotto la matrice di correlazione osservata, se il campione è estratto da una distribuzione normale multivariata. Le correlazioni sono pesate tramite l'inverso dell'unicità delle variabili. Viene utilizzato un algoritmo iterativo.
- **Fattorizzazione dell'asse principale.** Un metodo di estrazione dei fattori dalla matrice di correlazione originale con coefficienti di correlazione multipla al quadrato posti sulla diagonale come stime iniziali delle comunalità. Questi pesi di fattore vengono usati per stimare nuove comunalità che sostituiscono le vecchie stime sulla diagonale. Le iterazioni continuano fino a che le variazioni nelle comunalità da un'iterazione alla successiva soddisfano il criterio di convergenza per l'estrazione.

- **Alfa.** Un metodo per l'estrazione dei fattori che considera le variabili nell'analisi come un campione estratto dall'universo delle variabili potenziali. Questo metodo massimizza l'Alpha di Cronbach dei fattori.
- **Fattorizzazione immagine (Analisi fattoriale).** Metodo di estrazione dei fattori sviluppato da Guttman e basato sulla teoria dell'immagine. La parte comune della variabile, detta immagine parziale, viene definita come la sua regressione lineare sulle variabili rimanenti, piuttosto che in funzione di fattori ipotetici.

Analizza. Consente di specificare una matrice di correlazione o una matrice di covarianza.

- **Matrice di correlazione.** Può risultare utile se le variabili dell'analisi vengono misurate su scale diverse.
- **Matrice di covarianza.** Può risultare utile se si intende applicare l'analisi fattoriale a più gruppi con varianze diverse per ogni variabile.

Estrai. È possibile mantenere tutti i fattori con autovalori superiori al valore specificato oppure mantenere solo il numero di fattori specificato.

Visualizzazione. Consente di richiedere la soluzione fattoriale non ruotata e un grafico decrescente degli autovalori.

- **Soluzione non ruotata.** Visualizza la matrice dei pesi fattoriali, le comunalità e gli autovalori della soluzione fattoriale.
- **Grafico decrescente autovalori.** Un grafico della varianza associata a ciascun fattore. Questo grafico viene usato per determinare il numero di fattori da mantenere. In genere il grafico mostra una decisa diminuzione di pendenza nel punto in cui entrano in gioco i fattori meno rilevanti.

Massimo numero di iterazioni per la convergenza. Consente di specificare il numero massimo di passaggi che l'algoritmo può eseguire per valutare la soluzione.

Analisi fattoriale: Rotazione

Figura 34-5
Finestra di dialogo Analisi fattoriale: Rotazione

Metodo. Consente di selezionare il metodo di rotazione fattoriale. I metodi disponibili sono: varimax, equamax, quartimax, oblimin diretto e promax.

- **Varimax (Analisi fattoriale).** Un metodo di rotazione ortogonale che minimizza il numero di variabili che caricano fortemente ciascun fattore. Questo metodo semplifica l'interpretazione dei fattori.
- **Metodo oblimin diretto.** Un metodo di rotazione obliqua (non ortogonale). Quando delta vale 0 (il valore di default), le soluzioni sono per la maggior parte oblique. Quando delta diventa negativo e aumenta in valore assoluto, i fattori cominciano a essere meno obliqui. Inserire un numero minore o uguale a 0,8 per sovrascrivere il valore di default.
- **Metodo Quartimax.** Un metodo di rotazione che rende minimo il numero di fattori necessari per spiegare ogni variabile. Questo metodo semplifica l'interpretazione delle variabili osservate.
- **Equamax.** Un metodo di rotazione che rappresenta una combinazione tra il metodo varimax, che minimizza i fattori, e il metodo quartimax, che semplifica le variabili. È una combinazione dei metodi Varimax e Quartimax.
- **Rotazione Promax.** Una rotazione obliqua che ammette la correlazione fra fattori. Questa rotazione può essere utile per file di grandi dimensioni in quanto è più veloce del metodo Oblimin.

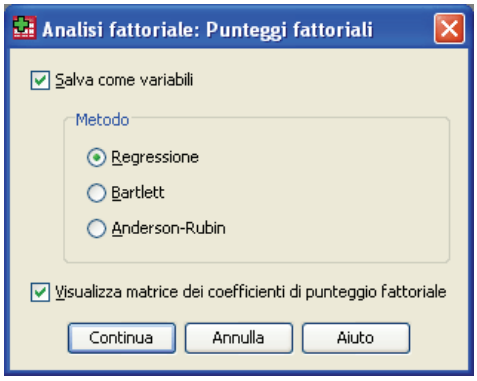
Visualizzazione. Consente di includere l'output nella soluzione ruotata, nonché i grafici dei pesi fattoriali per i primi due o tre fattori.

- **Soluzione ruotata (Analisi fattoriale).** Per ottenere la soluzione ruotata occorre aver selezionato un metodo di rotazione. Per le rotazioni ortogonali vengono visualizzate la matrice ruotata dei modelli e la matrice di trasformazione. Per le rotazioni oblique vengono visualizzate la matrice dei modelli, la matrice di struttura e la matrice di correlazione dei fattori.
- **Grafico dei pesi fattoriali.** Grafico tridimensionale dei pesi dei primi tre fattori. Per le soluzioni a due fattori viene prodotto un grafico bidimensionale. Assente se l'analisi estrae un solo fattore. Se è stata richiesta la rotazione, il grafico visualizza la soluzione ruotata.

Massimo numero di iterazioni per la convergenza. Consente di specificare il massimo numero di passaggi che l'algoritmo può eseguire per completare la rotazione.

Analisi fattoriale: Punteggi fattoriali

Figura 34-6
Finestra di dialogo Analisi fattoriale: Punteggi fattoriali



Salva come variabili. Consente di creare una nuova variabile per ciascun fattore nella soluzione finale.

Metodo. I metodi alternativi per il calcolo dei punteggi fattoriali sono regressione, Bartlett e Anderson-Rubin.

- **Metodo di regressione.** Un metodo per calcolare i coefficienti dei punteggi fattoriali. I punteggi prodotti hanno media 0 e varianza pari al quadrato della correlazione multipla fra i punteggi stimati e i valori reali dei fattori. I punteggi possono essere correlati anche quando i fattori sono ortogonali.
- **Punteggi di Bartlett.** Un metodo di stima dei coefficienti di punteggio fattoriale. I punteggi fattoriali di Bartlett hanno media pari a 0. La somma dei quadrati dei singoli fattori sull'intervallo delle variabili è minimizzata.
- **Metodo di Anderson-Rubin.** Un metodo per calcolare i coefficienti dei punteggi fattoriali; rappresenta una modifica del metodo di Bartlett per assicurare l'ortogonalità dei fattori stimati. I punteggi forniti hanno una media pari a 0, deviazione standard pari a 1 e risultano ortogonali (non correlati fra loro).

Visualizza matrice dei coefficienti di punteggio fattoriale. Mostra i coefficienti per cui vengono moltiplicate le variabili per ottenere i punteggi fattoriali. Vengono visualizzate anche le correlazioni tra i punteggi fattoriali.

Analisi fattoriale: Opzioni

Figura 34-7
Finestra di dialogo Analisi fattoriale: Opzioni

Valori mancanti. Consente di specificare le modalità di gestione dei valori mancanti. Le scelte disponibili sono: Esclusione **listwise**, Esclusione **pairwise** o Sostituisci con la media.

Formato visualizzazione coefficienti. Consente di controllare alcuni aspetti delle matrici di output. È possibile ordinare i coefficienti per dimensioni ed eliminare i coefficienti con valori assoluti inferiori al valore specificato.

Opzioni aggiuntive del comando FACTOR

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare i criteri di convergenza per l'iterazione durante l'estrazione e la rotazione.
- Specificare i singoli grafici dei fattori ruotati.
- Specificare quanti punteggi di fattori salvare.
- Specificare i valori diagonali per il metodo di calcolo dei fattori dell'asse principale.
- Scrivere le matrici di correlazione o le matrici dei fattori sul disco per poterle analizzare in seguito.
- Leggere e analizzare le matrici di correlazione o dei fattori.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
35

Scelta di una procedura per il raggruppamento

È possibile eseguire cluster analysis utilizzando le procedure Analisi Cluster TwoStep, Cluster gerarchica o Cluster con metodo delle k-medie. Ogni procedura utilizza un algoritmo diverso per la creazione di cluster e include opzioni non disponibili nelle altre procedure.

Analisi Cluster TwoStep. La procedura Analisi Cluster TwoStep risulta appropriata per molte applicazioni. Rende disponibili le seguenti funzioni univoche:

- Selezione automatica del numero di cluster più appropriato, oltre a misure per la scelta tra modelli di cluster.
- Possibilità di creare contemporaneamente due modelli di cluster basati su variabili categoriali e continue.
- Possibilità di salvare il modello di cluster in un file XML esterno e quindi di leggere tale file e aggiornare il modello di cluster utilizzando i dati più recenti.

Inoltre, questa procedura consente di analizzare file di dati di grandi dimensioni.

Cluster gerarchica. La procedura Cluster gerarchica è limitata a file di dati di dimensioni minori (ad esempio per il raggruppamento di centinaia di oggetti) e rende disponibili le seguenti funzioni univoche:

- Possibilità di raggruppare casi o variabili in cluster.
- Possibilità di calcolare un intervallo di soluzioni possibili e di salvare l'appartenenza a un cluster per ognuna di queste soluzioni.
- Diversi metodi per la formazione di cluster, la trasformazione delle variabili e la misurazione della dissimilarità tra cluster.

La procedura Cluster gerarchica analizza variabili per intervallo (continue), binarie o di conteggio, purché le variabili siano tutte dello stesso tipo.

Cluster con metodo delle k-medie. La procedura Cluster con metodo delle k-medie è limitata a dati continui e richiede che il numero di cluster venga specificato anticipatamente, rendendo tuttavia disponibili le seguenti funzioni univoche:

- Possibilità di salvare le distanze dai centri di cluster per ogni oggetto.
- Possibilità di leggere centri iniziali di cluster da un file SPSS esterno e salvare centri finali di cluster in un file esterno in formato SPSS Statistics.

Inoltre, questa procedura consente di analizzare file di dati di grandi dimensioni.

Capitolo
36

Analisi cluster TwoStep

L'analisi Cluster TwoStep è uno strumento di esplorazione che consente di rilevare raggruppamenti naturali, o cluster, all'interno di insiemi di dati, che non sarebbero altrimenti evidenti. L'algoritmo utilizzato da questa procedura presenta diverse caratteristiche che lo differenziano dalle tecniche di raggruppamento tradizionali:

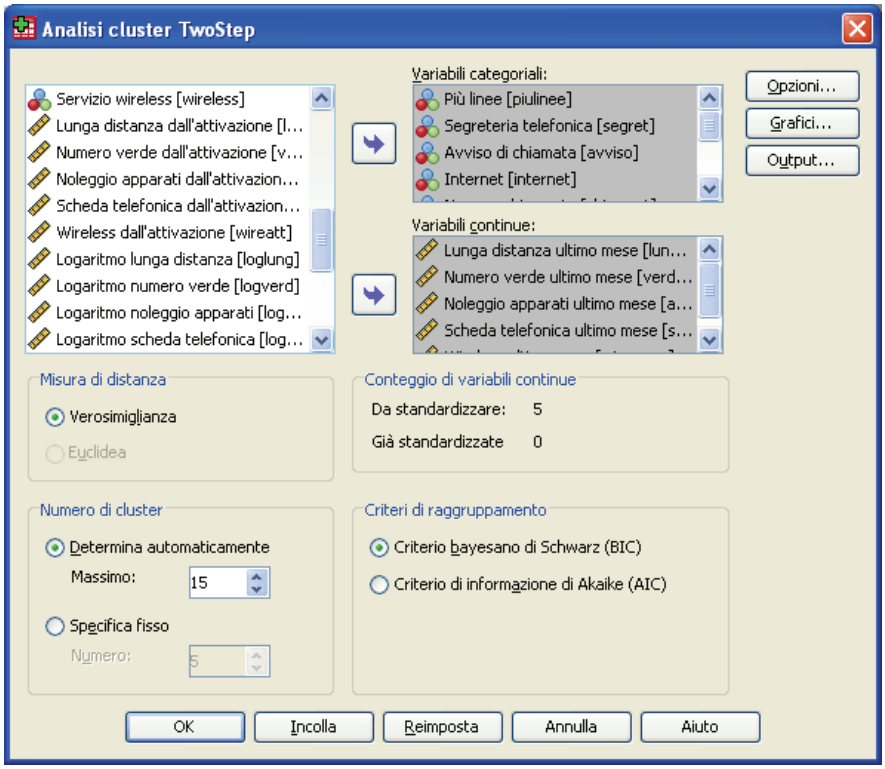
- **Gestione di variabili categoriali e continue.** Se le variabili sono indipendenti, è possibile applicare una distribuzione normale multinomiale congiunta alle variabili categoriali e continue.
- **Selezione automatica del numero di cluster.** Mediante il confronto tra i valori dei criteri di scelta di modello appartenenti a diverse soluzioni di raggruppamento, la procedura è in grado di determinare automaticamente il numero ottimale di cluster.
- **Scalabilità.** Mediante la creazione di un albero delle caratteristiche dei cluster (CF) che fornisce un riepilogo dei record, l'algoritmo TwoStep consente di analizzare file di dati di grandi dimensioni.

Esempio. I produttori di articoli al dettaglio e prodotti per i consumatori applicano regolarmente tecniche di raggruppamento ai dati relativi alle abitudini di acquisto dei propri clienti, al sesso, all'età, al livello di reddito e così via. In questo modo adattano le strategie di sviluppo del prodotto e di mercato ad ogni gruppo di consumatori al fine di aumentare le vendite e accrescere la fedeltà alla marca.

Statistiche. La procedura fornisce criteri di informazione (AIC o BIC) in base ai numeri di cluster nella soluzione, frequenze di cluster per il raggruppamento finale e statistiche descrittive in base al cluster per il raggruppamento finale.

Grafici. La procedura fornisce grafici a barre e grafici a torta per le frequenze dei cluster e grafici relativi all'importanza delle variabili.

Figura 36-1
Finestra di dialogo Analisi Cluster TwoStep



Misura di distanza. Questa selezione determina la modalità di calcolo della similarità tra due cluster.

- **Verosimiglianza.** La misura di verosimiglianza applica una distribuzione per probabilità alle variabili. Si suppone che le variabili continue vengano distribuite normalmente, mentre le variabili categoriali in base al modello multinomiale. Si suppone che tutte le variabili siano indipendenti.
- **Euclidea.** La misura euclidea è la distanza in "linea retta" tra due cluster. Può essere utilizzata solo quando tutte le variabili sono continue.

Numero di cluster. Questa selezione consente di specificare la modalità di definizione del numero dei cluster.

- **Determina automaticamente.** La procedura determina automaticamente il numero di cluster ottimale, mediante i criteri specificati nel gruppo Criteri di raggruppamento. È inoltre possibile immettere un numero intero positivo per definire il numero di cluster massimo che la procedura dovrà prendere in considerazione.
- **Specifica fisso.** Consente di definire un numero fisso di cluster nella soluzione. Immettere un numero intero positivo.

Conteggio di variabili continue. Questo gruppo fornisce un riepilogo delle opzioni di standardizzazione relative alle variabili continue specificate nella finestra di dialogo Opzioni. Per ulteriori informazioni, vedere Opzioni di Analisi Cluster TwoStep a pag. 431.

Criteri di raggruppamento. Questa selezione determina la modalità di definizione del numero dei cluster mediante l'algoritmo di raggruppamento automatico. È possibile specificare il modello Criterio bayesiano di Schwarz (BIC, Bayesian Information Criterion) o il modello Criterio di informazione di Akaike (AIC, Akaike Information Criterion).

Dati. Questa procedura può essere utilizzata sia con le variabili continue sia con le variabili categoriali. I casi rappresentano gli oggetti da raggruppare e le variabili corrispondono agli attributi in base ai quali viene eseguito il raggruppamento.

Ordine dei casi. La funzione Albero delle caratteristiche cluster e la soluzione finale possono dipendere dall'ordine dei casi. Per ridurre al minimo gli effetti dell'ordine, disporre i casi in ordine casuale. Può essere utile ottenere più soluzioni diverse con casi disposti in ordini casuali diversi per verificare la stabilità di una soluzione specifica. Nei casi in cui questa operazione è complessa a causa delle dimensioni eccessive dei file, è possibile effettuare più operazioni con un campione di casi disposti in ordini casuali diversi.

Assunzioni. La misura di distanza della verosimiglianza assume che le variabili nel modello di cluster siano indipendenti. A ogni variabile continua inoltre si suppone inoltre che sia associata una distribuzione normale o gaussiana, mentre a ogni variabile categoriale una distribuzione multinomiale. La verifica empirica interna indica che la procedura è piuttosto robusta rispetto alle violazioni sia delle assunzioni di indipendenza sia delle assunzioni di distribuzione, ma è consigliabile verificare fino a che punto tali assunzioni vengano soddisfatte.

Utilizzare la procedura Correlazioni bivariate per verificare l'indipendenza di due variabili continue. Utilizzare la procedura Tavole di contingenza per verificare l'indipendenza di due variabili categoriali. Utilizzare la procedura Medie per verificare l'indipendenza tra una variabile continua e una variabile categoriale. Utilizzare la procedura Esplora per verificare la normalità di una variabile continua. Utilizzare la procedura Test Chi-quadrato per verificare se per una variabile categoriale è stata specificata una distribuzione multinomiale.

Per ottenere un'analisi Cluster TwoStep

- Dai menu, scegliere:

  Analizza
  Classifica
  Cluster TwoStep...

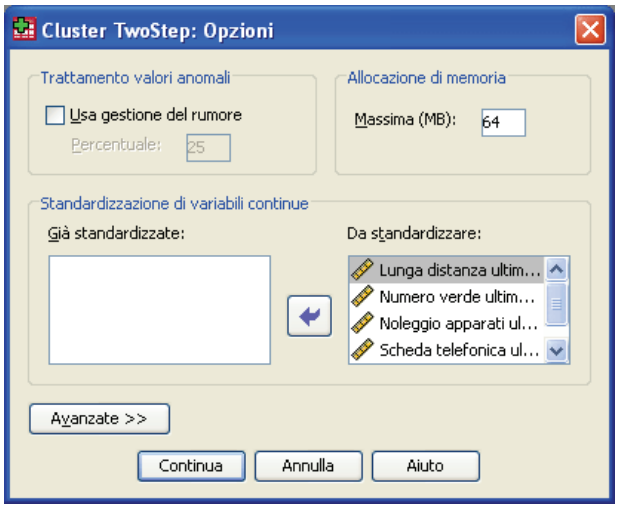
- Selezionare una o più variabili categoriali o continue.

Se lo si desidera, è possibile:

- Modificare i criteri in base ai quali sono stati creati i cluster.
- Selezionare le impostazioni di gestione del rumore, allocazione di memoria, standardizzazione delle variabili e input del modello di cluster.
- Richiedere tabelle e grafici facoltativi.
- Salvare i risultati del modello nel file di lavoro o in un file XML esterno.

Opzioni di Analisi Cluster TwoStep

Figura 36-2
Finestra di dialogo Cluster TwoStep: Opzioni



Trattamento valori anomali. Questo gruppo consente di trattare i valori anomali soprattutto durante il raggruppamento, se l'albero delle caratteristiche dei cluster risulta pieno. L'albero delle caratteristiche dei cluster è pieno quando non è più in grado di accettare casi in un nodo foglia e nessun nodo foglia può essere suddiviso.

- Se si seleziona la gestione del rumore e l'albero delle caratteristiche dei cluster risulta pieno, sarà possibile ampliarlo posizionando i casi mal distribuiti in più foglie all'interno di una foglia specifica per il "rumore". Una foglia contiene casi mal distribuiti quando il numero dei casi è inferiore alla percentuale specificata per la dimensione massima della foglia. Dopo aver ampliato la struttura, i valori anomali vengono inseriti nell'albero delle caratteristiche dei cluster, se possibile. Altrimenti, vengono eliminati.
- Se non si seleziona la gestione del rumore e l'albero delle caratteristiche dei cluster risulta pieno, sarà possibile ampliarlo utilizzando una soglia di modifica della distanza più elevata. Dopo il raggruppamento finale, i valori non assegnati a un cluster vengono definiti valori anomali. Al cluster dei valori anomali viene assegnato il numero di identificazione –1 e non viene incluso nel conteggio dei cluster.

Allocazione di memoria. Questo gruppo consente di specificare in megabyte (MB) la quantità massima di memoria che l'algoritmo del cluster può utilizzare. Se questa quantità massima viene superata, la procedura utilizzerà il disco per memorizzare le informazioni che non è possibile inserire nella memoria. Specificare un numero maggiore o uguale a 4.

- Per informazioni sul valore massimo per il sistema, rivolgersi all'amministratore del sistema.
- La ricerca dei cluster corretti o desiderati mediante l'algoritmo potrebbe non riuscire, se il valore è troppo basso.

Standardizzazione delle variabili. L'algoritmo di raggruppamento funziona con variabili continue standardizzate. Qualsiasi variabile non standardizzata deve essere impostata come "Da standardizzare". Per risparmiare tempo e calcoli, è possibile impostare le variabili continue già standardizzate come "Già standardizzate".

Opzioni avanzate

Criteri di ottimizzazione dell'albero delle caratteristiche (CF). Le seguenti impostazioni dell'algoritmo di raggruppamento riguardano in modo specifico l'albero delle caratteristiche dei cluster e devono essere modificate con cautela:

- **Soglia modifica distanza iniziale.** Si tratta della soglia iniziale utilizzata per ampliare l'albero delle caratteristiche dei cluster. Se dopo l'inserimento di un determinato caso in una foglia dell'albero delle caratteristiche dei cluster, la distanza risulta inferiore alla soglia, la foglia non viene suddivisa. Se la distanza supera la soglia, la foglia può essere suddivisa.
- **Ramificazioni massime (per nodo foglia).** Il numero massimo di nodi figlio per un nodo foglia.
- **Massima profondità struttura (livelli).** Il numero massimo di livelli dell'albero delle caratteristiche dei cluster.
- **Massimo numero di nodi possibile.** Indica il numero massimo dei nodi dell'albero delle caratteristiche dei cluster che può essere potenzialmente generato dalla procedura, in base alla funzione $(b^{d+1} - 1) / (b - 1)$, dove b sono le ramificazioni massime e d la profondità massima. Tenere presente che un albero delle caratteristiche dei cluster di dimensioni eccessive può rappresentare un peso considerevole per le risorse del sistema e compromettere quindi le prestazioni della procedura. Ogni nodo richiede un minimo di 16 byte.

Aggiornamento del modello gruppi. Questo gruppo consente di importare e aggiornare un modello di cluster generato da un'analisi precedente. Il file di input contiene l'albero delle caratteristiche dei cluster in formato XML. Il modello viene quindi aggiornato con i dati nel file attivo. È necessario selezionare i nomi delle variabili nella finestra di dialogo principale in base allo stesso ordine dell'analisi precedente. Il file XML non viene modificato, a meno che le nuove informazioni relative al modello non vengano inserite nello stesso file. Per ulteriori informazioni, vedere Output di Analisi Cluster TwoStep a pag. 434.

Se viene selezionato l'aggiornamento del modello dei cluster, verranno utilizzate le opzioni per la generazione dell'albero delle caratteristiche dei cluster specificate per il modello originale. Vengono quindi utilizzate le impostazioni del modello salvato relative a misura di distanza, gestione del rumore, allocazione di memoria e ottimizzazione dell'albero delle caratteristiche dei cluster, mentre qualsiasi nuova impostazione specificata nelle finestre di dialogo viene ignorata.

Nota: durante l'aggiornamento di un modello di cluster, la procedura si basa sul presupposto che nessuno dei casi selezionati nel file dati attivo sia stato utilizzato per creare il modello di cluster originale. La procedura si basa inoltre sul presupposto che i casi utilizzati nell'aggiornamento del modello di cluster provengono dalla stessa popolazione di casi utilizzati per creare il modello originale, le medie e le varianze delle variabili continue e i livelli delle variabili categoriali devono quindi essere le stesse per i due insiemi di casi. Se gli insiemi di casi precedenti e correnti provengono da una popolazione eterogenea, è necessario eseguire la procedura Analisi Cluster TwoStep negli insiemi di casi combinati per ottenere risultati ottimali.

Grafici di Analisi Cluster TwoStep

Figura 36-3
Finestra di dialogo Cluster TwoStep: Grafici

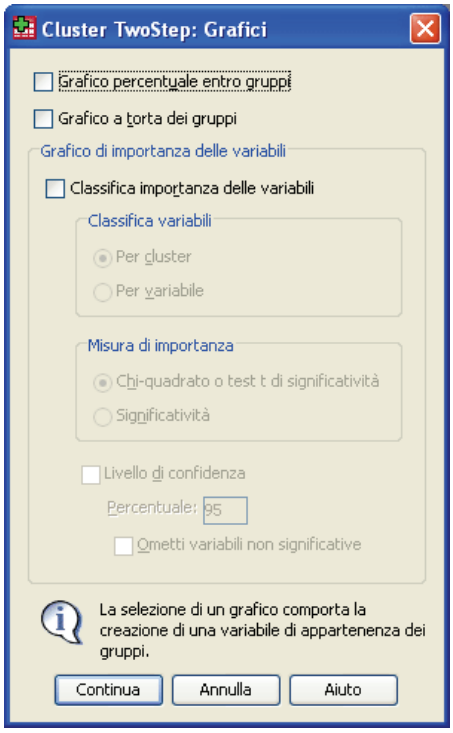


Grafico percentuale entro gruppi. Visualizza i grafici della variazione di ogni variabile all'interno di un cluster. Per ogni variabile categoriale viene creato un grafico a barre raggruppato, in cui viene illustrata la frequenza delle categorie in base all'ID cluster. Per ogni variabile continua viene invece creato un grafico degli errori, in cui vengono riportati gli errori in base all'ID cluster.

Grafico a torta dei gruppi. Visualizza un grafico a torta con le percentuali e i conteggi delle osservazioni all'interno di ogni cluster.

Grafico di importanza delle variabili. Visualizza diversi grafici in cui viene illustrata l'importanza di ogni variabile all'interno di ogni cluster. L'output viene ordinato in base al livello di importanza di ogni variabile.

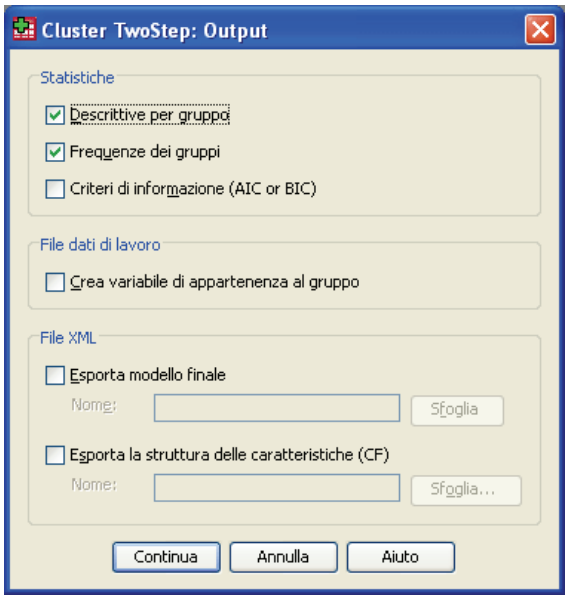
- **Classifica variabili.** Questa opzione consente di definire se i grafici verranno creati per ogni cluster (Per cluster) o per ogni variabile (Per variabile).
- **Misura di importanza.** Questa opzione consente di selezionare la misura di importanza della variabile che verrà utilizzata per creare il grafico. Chi-quadrato o test t di significatività segnala una statistica chi-quadrato di Pearson come importanza della variabile categoriale e una statistica *t* come importanza della variabile continua. Significatività segnala il valore *p* meno uno per il test di uguaglianza delle medie di una variabile continua e la frequenza attesa per l'insieme di dati globale di una variabile categoriale.
- **Livello di confidenza.** Questa opzione consente di impostare il livello di confidenza per il test di uguaglianza della distribuzione di una variabile all'interno di un cluster rispetto alla distribuzione globale della variabile. Specificare un numero minore di 100 e maggiore o

uguale a 50. Il valore del livello di confidenza è rappresentato da una riga verticale nei grafici di importanza delle variabili, se i grafici vengono creati in base alla variabile o se la misura di significatività è riportata nel grafico.

- **Ometti variabili non significative.** Le variabili non significative in base al livello di confidenza specificato non vengono visualizzate nei grafici di importanza delle variabili.

Output di Analisi Cluster TwoStep

Figura 36-4
Finestra di dialogo Cluster TwoStep: Output



Statistiche. Questo gruppo fornisce le opzioni per la visualizzazione delle tabelle dei risultati dei raggruppamenti. Le statistiche descrittive e le frequenze dei cluster vengono fornite per il modello di cluster finale, mentre la tabella dei criteri delle informazioni contiene i risultati di un intervallo di soluzioni di cluster.

- **Descrittive per gruppo.** Visualizza due tabelle che descrivono le variabili presenti in ogni cluster. In una tabella, le medie e le deviazioni standard vengono segnalate per le variabili continue in base al cluster. L'altra tabella riporta le frequenze delle variabili categoriali per cluster.
- **Frequenze dei gruppi.** Visualizza una tabella che riporta il numero di osservazioni in ogni cluster.
- **Criteri di informazione (AIC or BIC).** Visualizza una tabella che include i valori di AIC o BIC, in base ai criteri scelti nella finestra di dialogo principale, per i diversi numeri di cluster. Questa tabella viene fornita solo quando il numero di cluster viene determinato in modo automatico. Se il numero di cluster è fisso, questa impostazione viene ignorata e la tabella non viene fornita.

File dati di lavoro. Questo gruppo consente di salvare le variabili all'interno del file dati attivo.

- **Crea variabile di appartenenza al gruppo.** Questa variabile contiene un numero di identificazione del cluster per ogni caso. Il nome di questa variabile è *tsc_n*, dove *n* è un numero intero positivo che indica l'ordinale dell'operazione di salvataggio del file dati attivo eseguita mediante questa procedura in una determinata sessione.

File XML. Il modello di cluster finale e l'albero delle caratteristiche dei cluster rappresentano due tipi di file di output che possono essere esportati in formato XML.

- **Esporta modello finale.** Il modello di cluster finale viene esportato nel file specificato in formato XML (PMML). SmartScore e la versione del server di SPSS Statistics (un prodotto separato) possono utilizzare questo file del modello per applicare le informazioni del modello ad altri file dati a scopo di assegnazione dei punteggi.
- **Esporta l'albero delle caratteristiche (CF).** Questa opzione consente di salvare lo stato corrente dell'albero delle caratteristiche dei cluster e aggiornarlo in un secondo momento utilizzando dati più aggiornati.

Capitolo
37

Cluster gerarchica

Questa procedura consente di identificare gruppi di casi relativamente omogenei in base alle caratteristiche selezionate, utilizzando un algoritmo che inizia con ciascun caso (o variabile) in un cluster distinto e che combina i cluster fino a quando ne rimane solo uno. È possibile analizzare le variabili semplici oppure scegliere una delle trasformazioni di standardizzazione disponibili. Le misure di similarità e dissimilarità vengono generate dalla procedura Distanze. A ciascun livello verranno visualizzate statistiche in base alle quali selezionare la soluzione migliore.

Esempio. Esistono gruppi di trasmissioni televisive identificabili che attraggono tipi di audience analoghi all'interno di ciascun gruppo? Utilizzando la cluster gerarchica è possibile raggruppare le trasmissioni televisive (casi) in gruppi omogenei in base alle caratteristiche degli spettatori. Questo metodo può essere utilizzato per identificare i segmenti di mercato. In alternativa, è possibile raggruppare le città (casi) in gruppi omogenei in modo che da poter selezionare città con caratteristiche confrontabili per verificare diverse strategie di mercato.

Statistiche. Programma di agglomerazione, matrice delle distanze (o similarità) e cluster di appartenenza per un'unica soluzione o una serie di soluzioni. Grafici: dendrogrammi e grafici a stalattite

Dati. Le variabili possono essere quantitative, binarie o dati di conteggio. Lo scaling delle variabili è molto importante in quanto le differenze di scaling possono influire sulle soluzioni dei cluster. Se lo scaling delle variabili presenta differenze notevoli (ad esempio, una variabile viene misurata in dollari e l'altra in anni), è consigliabile standardizzarle. Ciò può essere effettuato in modo automatico mediante la procedura Cluster gerarchica.

Ordine dei casi. Se le distanze assegnate o le similarità sono presenti nei dati iniziali o nei cluster aggiornati durante l'unione, la soluzione del cluster risultante può essere influenzata dall'ordine dei casi del file. Può essere utile ottenere più soluzioni diverse con casi disposti in ordini casuali diversi per verificare la stabilità di una soluzione specifica.

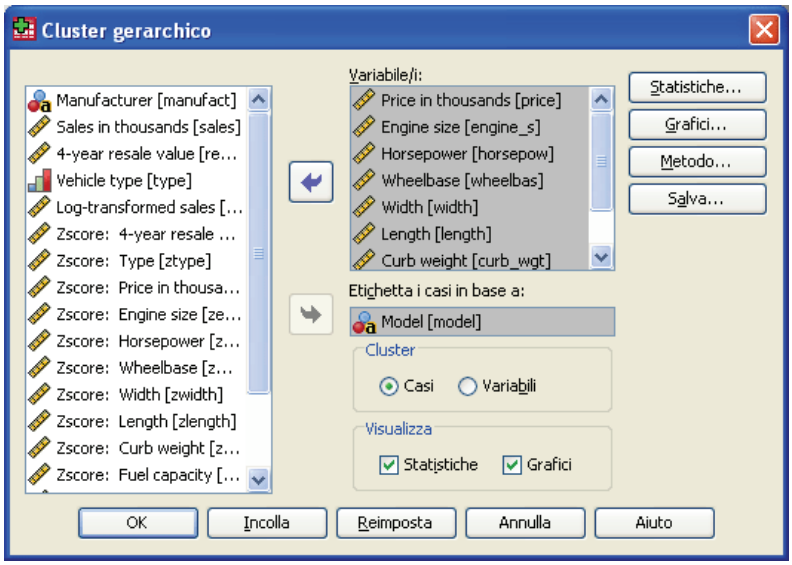
Assunzioni. Le misure di dissimilarità o di similarità utilizzate devono essere idonee per i dati analizzati. Per ulteriori informazioni sulla scelta delle misure di dissimilarità e similarità, vedere la procedura Distanze. È inoltre necessario includere nell'analisi tutte le variabili significative. L'omissione di variabili importanti può portare a soluzioni improprie. Poiché la cluster gerarchica rappresenta un metodo esplorativo, i risultati devono essere considerati provvisori finché non vengano confermati da un campione indipendente.

Per ottenere una cluster gerarchica

▶ Dai menu, scegliere:

Analizza
  Classificazione
    Cluster gerarchica...

Figura 37-1
Finestra di dialogo Cluster gerarchica

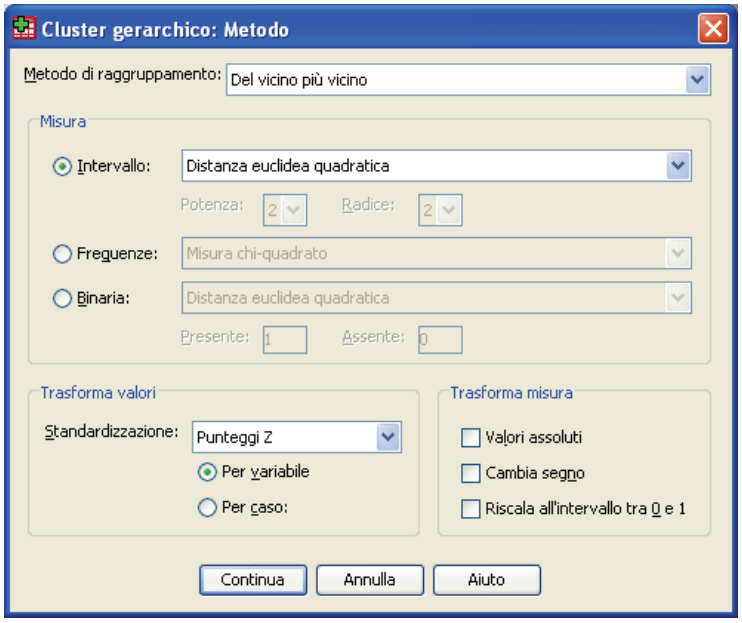


▶ Per raggruppare i casi in cluster è necessario selezionare almeno una variabile numerica. Per raggruppare le variabili in cluster è necessario selezionare almeno tre variabili numeriche.

È inoltre possibile selezionare una variabile di identificazione per etichettare i casi.

Cluster gerarchica: Metodo

Figura 37-2
Finestra di dialogo Cluster gerarchica: Metodo



Metodo di raggruppamento. Le alternative disponibili sono: Legame medio fra i gruppi, Legame medio entro gruppi, Del vicino più vicino, Del vicino più lontano, Centroide, Mediana e Ward.

Misura. Consente di specificare la misura di similarità o dissimilarità da utilizzare per il raggruppamento. Selezionare il tipo di dati e la misura di similarità o dissimilarità desiderata:

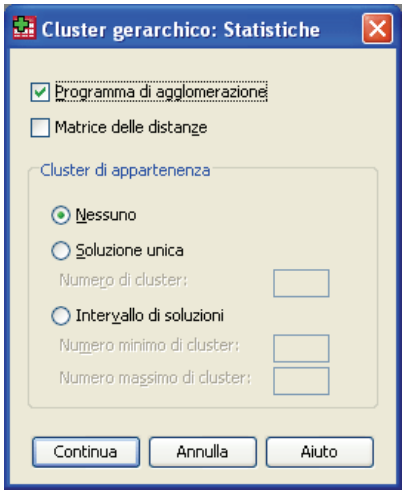
- **Intervallo.** Le alternative disponibili sono: Distanza euclidea, Distanza euclidea quadratica, Coseno, Correlazione di Pearson, Chebychev, City-Block, Minkowski e Personalizzato.
- **Frequenze.** Le alternative disponibili sono: Misura chi-quadrato e Misura phi-quadrato.
- **Binaria.** Le alternative disponibili sono: Distanza euclidea, Distanza euclidea quadratica, Differenza di dimensione, Differenza di modello, Varianza, Dispersione, Forma, Corrispondenza semplice, Correlazione phi a 4 punti, Lambda, *D* di Anderberg, Dice, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Lance e Williams, Ochiai, Rogers e Tanimoto, Russel e Rao, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Sokal e Sneath 4, Sokal e Sneath 5, *Y* di Yule e *Q*di Yule.

Trasforma valori. Consente di standardizzare i valori dei dati per casi o valori prima di calcolare le similarità (non disponibile per i dati binari). I metodi di standardizzazione disponibili sono: punteggi *z*, intervallo da −1 a 1, intervallo da 0 a 1, ampiezza massima di 1, media di 1 e deviazione standard di 1.

Trasforma misure. Consente di trasformare i valori generati dalla misura di distanza. Questi verranno applicati dopo il calcolo della misura di distanza. Le alternative disponibili sono: Valori assoluti, Cambia segno e Riscala all'intervallo 0–1.

Cluster gerarchica: Statistiche

Figura 37-3
Finestra di dialogo Cluster gerarchica: Statistiche



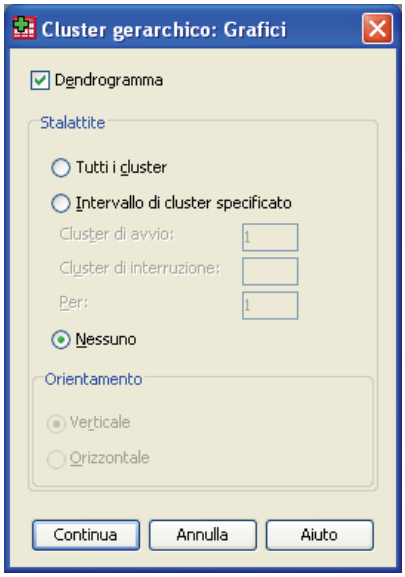
Programma di agglomerazione. Consente di visualizzare i casi o i cluster accorpati ad ogni stadio, le distanze tra i casi o i cluster da accorpare e l'ultimo livello di cluster in cui un caso (o una variabile) è stato accorpato al cluster.

Matrice delle distanze. Fornisce le distanze o le similarità tra gli elementi.

Cluster di appartenenza. Viene visualizzato il cluster a cui viene assegnato ciascun caso a uno o più stadi della combinazione dei cluster. Le opzioni disponibili sono Soluzione unica e Intervallo di soluzioni.

Cluster gerarchica: Grafici

Figura 37-4
Finestra di dialogo Cluster gerarchica: Grafici

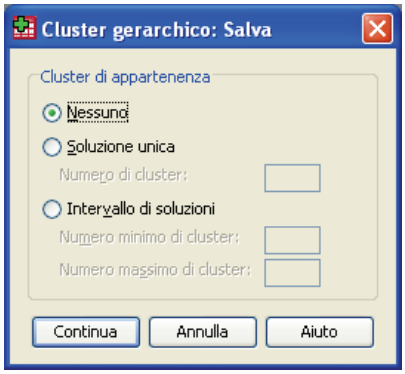


Dendrogramma. Viene visualizzato un **dendrogramma**. Utilizzando i dendrogrammi è possibile valutare la coesione dei cluster formati ed ottenere informazioni sul numero di cluster che è opportuno tenere.

A stalattite. Visualizza un **grafico a stalattite**, inclusi tutti i cluster o l'intervallo di cluster specificato. Nei grafici a stalattite vengono visualizzate informazioni sulle modalità con cui i casi vengono combinati in cluster ad ogni iterazione dell'analisi. Specificando l'orientamento desiderato è possibile selezionare un grafico verticale o orizzontale.

Cluster gerarchica: Salva nuove variabili

Figura 37-5
Finestra di dialogo Cluster gerarchica: Salva



Cluster di appartenenza. Consente di salvare i cluster di appartenenza per una soluzione unica o per un intervallo di soluzioni. Le variabili salvate possono essere utilizzate in analisi successive per valutare altre differenze tra i gruppi.

Funzioni aggiuntive della sintassi del comando CLUSTER

La procedura Cluster gerarchica usa la sintassi del comando `CLUSTER`. Il linguaggio della sintassi dei comandi consente inoltre di:

- Usare più metodi di raggruppamento in una singola analisi.
- Leggere ed analizzare una matrice di prossimità.
- Scrivere una matrice di prossimità sul disco per analizzarla in seguito.
- Specificare i valori per la potenza e la radice nella misura della distanza personalizzata (potenza).
- Specificare i nomi delle variabili salvate.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
38

Cluster con metodo delle k-medie

Questa procedura consente di identificare gruppi di casi relativamente omogenei in base alle caratteristiche selezionate, utilizzando un algoritmo in grado di gestire un elevato numero di casi. Tale algoritmo, tuttavia, richiede l'indicazione del numero di cluster. È possibile specificare i centri iniziali del cluster, se si conosce questa informazione. È possibile selezionare uno dei due metodi disponibili per la classificazione dei casi, ovvero l'aggiornamento iterativo dei centri cluster oppure la semplice classificazione. È possibile salvare l'appartenenza al cluster, le informazioni sulla distanza e i centri del cluster finali. È inoltre possibile specificare una variabile i cui valori possono essere utilizzati per etichettare l'output caso per caso. i può inoltre richiedere l'analisi delle statistiche *F* di varianza. Se da un lato queste statistiche sono opportunistiche, ovvero vengono eseguiti tentativi di raggruppamenti che presentino differenze, le corrispondenti dimensioni relative forniscono informazioni sul contributo apportato da ciascuna variabile alla separazione dei gruppi.

Esempio. Quali sono i gruppi di show televisivi che attraggono un pubblico analogo all'interno di ciascun gruppo? Il metodo cluster *k*-medie consente di raggruppare gli show televisivi (casi) in*k* gruppi omogenei in base alle caratteristiche degli spettatori. Questo processo può essere utilizzato per identificare i segmenti di mercato. In alternativa, è possibile raggruppare le città (casi) in gruppi omogenei in modo che da poter selezionare città con caratteristiche confrontabili per verificare diverse strategie di mercato.

Statistiche. Soluzione completa: centri iniziali del cluster, tabella ANOVA. Per ciascun caso: informazioni sui cluster, distanza dal centro del cluster.

Dati. Le variabili devono essere quantitative a livello di intervallo o di rapporto. Se le variabili sono binarie o conteggi, utilizzare la procedura Cluster gerarchica.

Ordine dei casi e dei centri di cluster iniziale. L'algoritmo predefinito per la scelta dei centri di cluster iniziali varia a seconda dell'ordine dei casi. L'opzioneUsa medie mobili della finestra di dialogo Iterazioni rende la soluzione risultante potenzialmente dipendente dall'ordine dei casi, indipendentemente dai centri di cluster scelti inizialmente. Se si utilizza uno di questi metodi, può essere utile ottenere più soluzioni diverse con casi disposti in ordini casuali diversi per verificare la stabilità di una soluzione specifica. Per evitare problemi con l'ordine dei casi, è consigliabile specificare i centri di cluster iniziali ed evitare di usare l'opzione Usa medie mobili. Tuttavia, l'ordinamento dei centri di cluster iniziali può influire sulla soluzione se esistono distanze assegnate dai casi ai centri di cluster. Per valutare la stabilità di una soluzione, è possibile confrontare i risultati delle analisi con diverse permutazioni dei valori dei centri iniziali.

Assunzioni. Le distanze vengono calcolate utilizzando la distanza euclidea semplice. Se si desidera utilizzare un'altra misura di distanza o di similarità, utilizzare la procedura Cluster gerarchica. La scalatura delle variabili è un'operazione che deve essere effettuata con molta attenzione. Se le variabili vengono misurate con scale diverse (ad esempio se una variabile è espressa in dollari

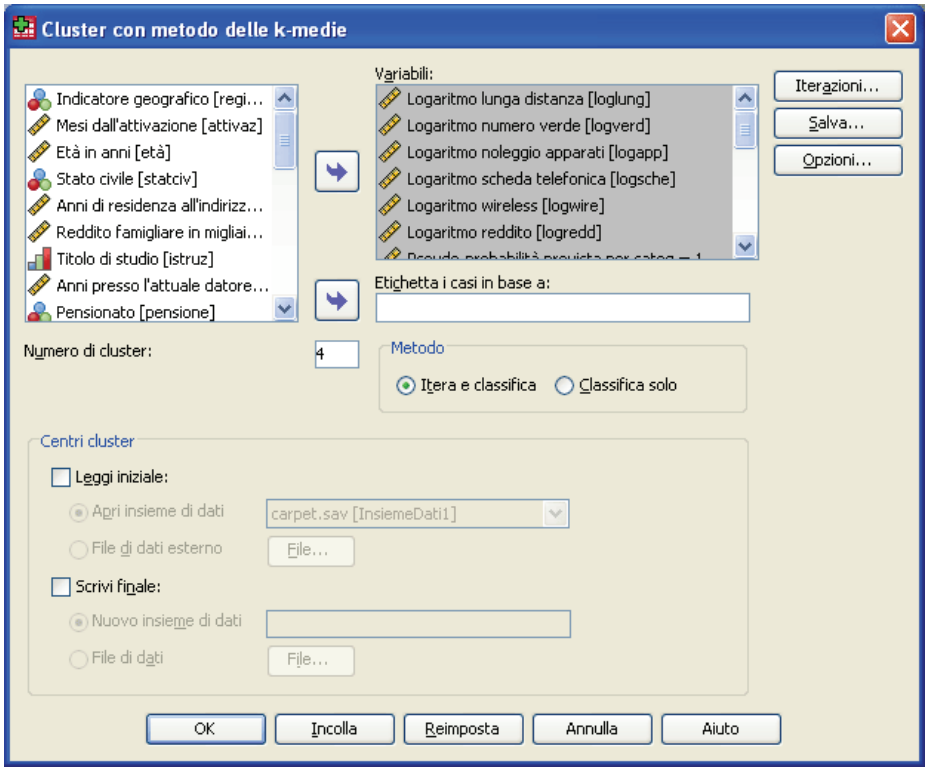
e un'altra è espressa in anni), i risultati possono essere fuorvianti. In questi casi è consigliabile standardizzare le variabili prima di procedere con l'analisi cluster *k*medie (utilizzando la procedura Descrittive). Questa procedura presume che sia stato selezionato il numero esatto di cluster e che siano state incluse tutte le variabili rilevanti. Se è stato selezionato un numero di cluster inesatto o sono state omesse variabili importanti, i risultati possono essere inattendibili.

Per ottenere un'analisi cluster K-medie

- Dai menu, scegliere:

Analizza
  Classifica
    Cluster k-medie...

Figura 38-1
Finestra di dialogo Cluster K-medie



- Selezionare le variabili da utilizzare nell'analisi cluster.

- Specificare il numero di cluster. Il numero di cluster specificato deve essere almeno di 2 e non deve essere maggiore al numero di casi del file dati.

- Selezionare il metodo Itera e classifica oppure il metodo Classifica soltanto.

- In alternativa, selezionare una variabile di identificazione per etichettare i casi.

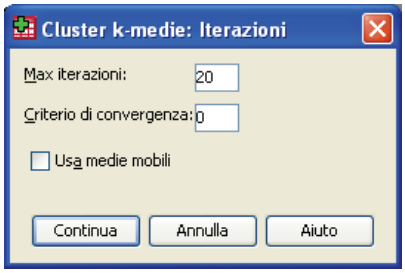
Efficienza dell'analisi cluster K-medie

Il comando Cluster *k*-medie è efficace principalmente in quanto non calcola le distanze tra tutte le coppie di casi, a differenza di numerosi algoritmi di raggruppamento, ad esempio quello utilizzato dal comando per la Cluster gerarchica di SPSS.

Per ottenere la massima efficienza, creare un campione di casi e utilizzare il metodo Itera e classifica per determinare i centri cluster. Selezionare Scrivi valori finali su file. Quindi, ripristinare tutto il file di dati e selezionare Classifica soltanto come metodo e selezionare Leggi valori iniziali per classificare tutto il file utilizzando i centri valutati per il campione. È possibile leggere o scrivere da un file o file di dati. I file di dati possono anche essere riutilizzati nella stessa sessione, ma non vengono salvati come file a meno che siano stati salvati come tali alla fine della sessione. I nomi degli insiemi di dati devono essere conformi alle regole di denominazione delle variabili. Per ulteriori informazioni, vedere Nomi delle variabili in Capitolo 5 a pag. 79.

Cluster K-medie: Iterazioni

Figura 38-2
Finestra di dialogo Cluster K-medie: Iterazioni



Nota: queste opzioni sono disponibili solo se si seleziona il metodo Itera e classifica nella finestra di dialogo Cluster con metodo delle K-medie.

Massimo numero di iterazioni. Consente di impostare il numero massimo di iterazioni per l'algoritmo *k*-medie. Le iterazioni si interromperanno al numero impostato, anche se il criterio di convergenza non viene soddisfatto. Il numero deve essere compreso tra 1 e 999.

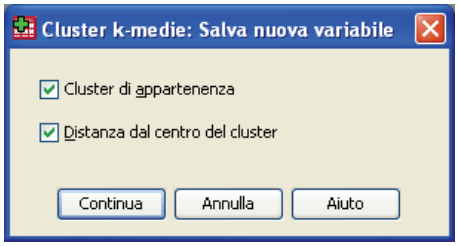
Per riprodurre l'algoritmo utilizzato dal comando Quick Cluster delle versioni di SPSS precedenti alla 5.0, impostare l'opzione Massimo numero di iterazioni su 1.

Criterio di convergenza. Determina il termine dell'iterazione. Rappresenta una proporzione della distanza minima fra i centri iniziali del cluster in modo che sia maggiore di 0 e minore di 1. Se, ad esempio, il criterio è 0,02, il processo di iterazione terminerà quando un'iterazione completa non è in grado di spostare i centri cluster di una distanza maggiore del 2% della distanza minima fra i centri iniziali del cluster.

Usa medie mobili. Consente di richiedere l'aggiornamento dei centri cluster in seguito all'assegnazione di ciascun caso Se non viene selezionata questa opzione, i nuovi centri del cluster verranno calcolati quando tutti i casi saranno stati assegnati.

Cluster K-medie: Salva

Figura 38-3
Finestra di dialogo Cluster K-medie: Salva nuove variabili



È possibile salvare informazioni sulla soluzione come nuove variabili da utilizzare in analisi successive:

Cluster di appartenenza. Consente di creare una nuova variabile che indica l'appartenenza finale al cluster di ciascun caso. I valori della nuova variabile sono compresi tra 1 e il numero di cluster.

Distanza dal centro del cluster. Consente di creare una nuova variabile che indica la distanza euclidea tra ciascun caso e il relativo centro di classificazione.

Cluster K-medie: Opzioni

Figura 38-4
Finestra di dialogo Cluster K-medie: Opzioni

Statistiche. È possibile selezionare le seguenti statistiche: centri iniziali del cluster, tabella ANOVA e informazioni sui cluster per ciascun caso.

- **Centri iniziali del cluster.** Prima stima delle medie delle variabili per ciascun cluster. In mancanza di indicazioni particolari, viene selezionato dai dati un numero di casi ben distanziati uguale al numero dei cluster. I centri dei cluster iniziali vengono usati per un primo ciclo di classificazione e poi vengono aggiornati.

- **Tabella ANOVA.** Visualizza una tabella di analisi della varianza con test F per ogni variabile di raggruppamento. I test F sono descrittivi e il livello di significatività fornisce informazioni utili. La tabella non viene creata se tutti i casi vengono assegnati a un solo cluster.
- **Cluster per ogni caso.** Visualizza per ogni caso il cluster di appartenenza finale e la distanza euclidea dal centro del cluster utilizzato per classificare il caso. Visualizza inoltre la distanza euclidea fra i centri finali.

Valori mancanti. Le opzioni disponibili sono Escludi casi listwise o Escludi casi pairwise.

- **Esclusione listwise.** Consente di escludere i casi coni valori mancanti per le variabili di raggruppamento dall'analisi.
- **Esclusione coppie.** Consente di assegnare i casi ai cluster in base alle distanze calcolate da tutte le variabili con valori non mancanti.

Opzioni aggiuntive del comando QUICK CLUSTER

La procedura Cluster K-medie usa la sintassi del comando `QUICK CLUSTER`. Il linguaggio della sintassi dei comandi consente inoltre di:

- Accettare i primi *k* casi come centri dei cluster iniziali per evitare di dover leggere i dati normalmente usati per stimarli.
- Specificare i centri iniziali dei cluster direttamente come parte della sintassi del comando.
- Specificare i nomi delle variabili salvate.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
39

Test non parametrici

La procedura Test non parametrici è costituita da diversi test che non richiedono ipotesi sulla forma della distribuzione sottostante:

Test Chi-quadrato. Consente di analizzare una variabile in categorie e di ottenere una statistica del chi-quadrato in base alle differenze tra frequenze osservate e attese.

Test binomiale. Consente di confrontare la frequenza osservata in ciascuna categoria di una variabile dicotomica con le frequenze attese dalla distribuzione binomiale.

Test delle successioni. Consente di verificare se l'ordine di occorrenza di due valori di una variabile è casuale.

Test di Kolmogorov-Smirnov per un campione. Consente di confrontare la funzione di distribuzione cumulativa osservata per una variabile con la distribuzione teorica specificata, che può essere normale, uniforme esponenziale o di Poisson.

Test per due campioni indipendenti. Consente di confrontare due gruppi di casi in base a una sola variabile. Sono disponibili il test *U* di Mann-Whitney, il test di Kolmogorov-Smirnov per due campioni, il test delle reazioni estreme di Moses e il test delle successioni di Wald-Wolfowitz.

Test per due campioni dipendenti. Consente di confrontare le distribuzioni di due variabili. Sono disponibili il test di Wilcoxon, il test del segno e il test di McNemar.

Test per diversi campioni indipendenti. Consente di confrontare due o più gruppi di casi in base alla stessa variabile. Sono disponibili il test di Kruskal-Wallis, il test della mediana e il test di Jonckheere-Terpstra.

Test per diversi campioni dipendenti. Consente di confrontare le distribuzioni di due o più variabili. Sono disponibili il test di Friedman, il test *W* di Kendall e il test *Q* di Cochran.

Per tutti i test precedentemente citati sono disponibili quartili e media, deviazione standard, valore minimo e massimo e numero di casi non mancanti.

Test Chi-quadrato

La procedura Test chi-quadrato permette di analizzare una variabile in categorie e di calcolare una statistica chi-quadrato. Questo test sulla bontà di adattamento permette di confrontare le frequenze osservate e attese in ciascuna categoria per verificare se tutte le categorie includono la stessa proporzione di valori o se includono una proporzione di valori specificati dall'utente.

Esempi. Il test chi-quadrato può essere utilizzato per determinare se in un sacchetto di gelatine di frutta è presente la stessa proporzione di blu, marrone, arancio, rosso e giallo. È inoltre possibile determinare se il sacchetto di gelatine contiene il 5% di blu, il 30% di marrone, il 10% di verde, il 20% di arancio, il 15% di rosso e il 15% di giallo.

Statistiche. Media, deviazione standard, valore minimo e massimo e quartili. Il numero e la percentuale di casi mancanti e non mancanti, il numero di casi osservati e attesi per ciascuna categoria, i residui e la statistica chi-quadrato.

Dati. Utilizzare variabili categoriali numeriche ordinate o non ordinate (livelli di misurazione ordinali o nominali). Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma.

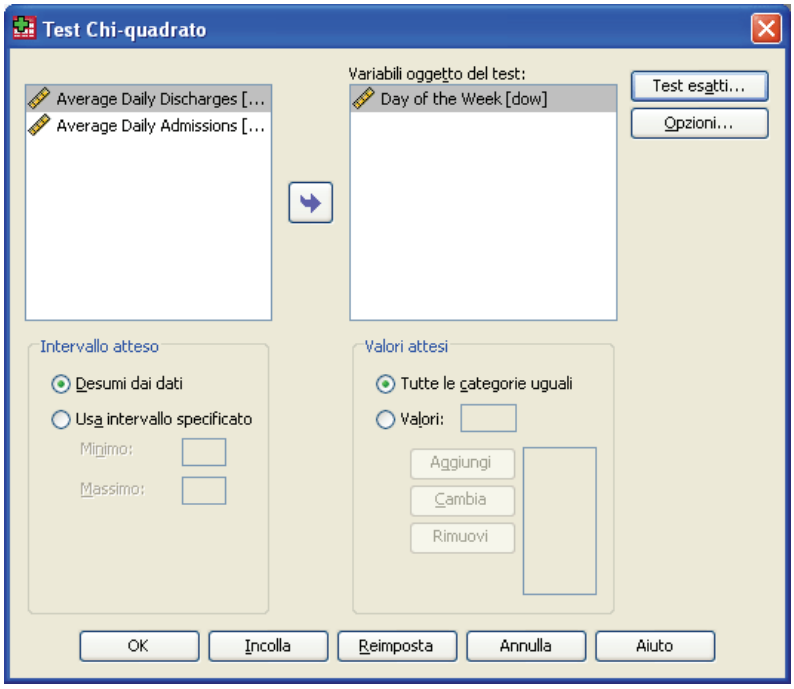
Assunzioni. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Si presume che i dati rappresentino un campione casuale. Le frequenze attese per ciascuna categoria devono essere come minimo pari a 1. Non più del 20% delle categorie possono avere frequenze attese inferiori a 5.

Per ottenere un test chi-quadrato

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    Chi-Quadrato...

Figura 39-1
Finestra di dialogo Test chi-quadrato



▶ Selezionare una o più variabili per il test. Ogni variabile produce un test distinto.

- È possibile fare clic su Opzioni per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

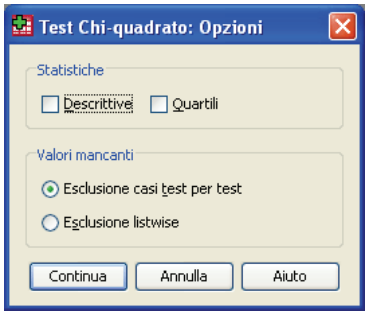
Intervallo e valori attesi del test chi-quadrato

Intervallo atteso. Per impostazione predefinita, ogni singolo valore della variabile è definito come categoria. Per definire categorie all'interno di un intervallo specifico, selezionare Usaintervallo specificato e inserire valori interi per il limite inferiore e superiore. Le categorie verranno definite come valori inclusi nell'intervallo specificato. I casi con valori al di fuori del minimo e del massimo specificato saranno esclusi dal test. Se, ad esempio, si specifica 1 per il limite inferiore e 4 per il limite superiore, per il test chi-quadrato verranno utilizzati solo i valori interi compresi tra 1 e 4.

Valori attesi. Per impostazione predefinita, tutte le categorie hanno valori attesi uguali. Per le categorie sono previste proporzioni attese definite dall'utente. Selezionare Valori, specificare un valore maggiore di 0 per ogni categoria della variabile del test e quindi fare clic su Aggiungi. I valori vengono elencati in ordine di inserimento, L'ordine dei valori è importante in quanto corrisponde all'ordine crescente dei valori delle categorie della variabile oggetto del test. Il primo valore della lista corrisponde al valore di gruppo minore della variabile, mentre l'ultimo corrisponde al valore maggiore. Gli elementi della lista dei valori vengono sommati e quindi ciascun valore viene diviso per la somma risultante per calcolare la proporzione di casi attesi nella categoria corrispondente. Ad esempio, una lista valori formata da 3, 4, 5, 4 specifica le proporzioni attese 3/16, 4/16, 5/16 e 4/16.

Test chi-quadrato: Opzioni

Figura 39-2
Finestra di dialogo Test chi-quadrato: Opzioni



Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (test chi-quadrato)

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare valori minimi e massimi diversi o frequenze attese diverse per variabili diverse (con il sottocomando `CHISQUARE`).
- Eseguire il test confrontando la stessa variabile con diverse frequenze attese o utilizzando diversi intervalli (con il sottocomando `EXPECTED`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Test binomiale

Grazie alla procedura del test binomiale è possibile confrontare le frequenze osservate delle due categorie di una variabile dicotomica con le frequenze previste in presenza di una distribuzione binomiale con il parametro di probabilità specificato. Per impostazione predefinita, il parametro di probabilità per entrambi i gruppi è 0,5. Per modificare le probabilità, è possibile inserire una proporzione di prova per il primo gruppo. La probabilità per il secondo gruppo sarà uguale a 1 meno la probabilità specificata per il primo gruppo.

Esempio. Quando si lancia in aria una moneta, la probabilità che esca testa è pari a 1/2. In base a questa ipotesi, la moneta viene lanciata in aria 40 volte e i risultati (testa o croce) vengono registrati. Dal test binomiale può risultare che per i 3/4 dei lanci della moneta è uscita testa e che il livello di significatività è molto basso (0,0027). Questi risultati indicano che la probabilità che venga testa molto spesso non è pari a 1/2; pertanto, la stima sul comportamento della moneta probabilmente risulta distorta.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili.

Dati. Le variabili incluse nel test devono essere numeriche e dicotomiche. Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma. Una **variabile dicotomica** è una variabile che prevede solo due possibili valori: *Sì* o *No*, *Vero* o *Falso*, 0 o 1 e così via. Il primo valore rilevato nell'insieme di dati definisce il primo gruppo e l'altro valore definisce il secondo gruppo. Se le variabili sono dicotomiche, è necessario specificare un punto di divisione. Utilizzando il punto di divisione è possibile assegnare al primo gruppo i casi con valori inferiori o uguali al punto di divisione e i rimanenti casi al secondo gruppo.

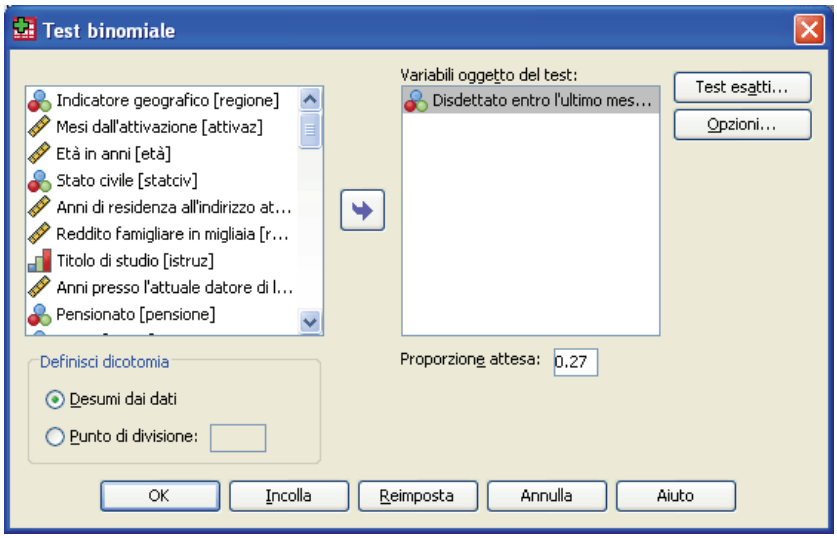
Assunzioni. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Si presume che i dati rappresentino un campione casuale.

Per ottenere un test binomiale

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    Binomiale...

Figura 39-3
Finestra di dialogo Test binomiale

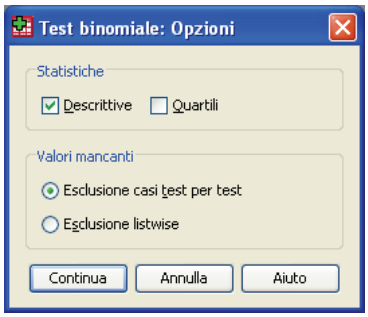


▶ Selezionare una o più variabili numeriche oggetto del test.

▶ È possibile fare clic su Opzioni per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Test binomiale: Opzioni

Figura 39-4
Finestra di dialogo Test binomiale: Opzioni



Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile da verificare verranno esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TEST (test binomiale)

Il linguaggio della sintassi dei comandi consente inoltre di:

- Selezionare gruppi specifici (ed escluderne altri) quando una variabile ha più di due categorie (con il sottocomando `BINOMIAL`).
- Specificare diversi punti di divisione o probabilità per variabili diverse (con il sottocomando `BINOMIAL`).
- Eseguire test confrontando la stessa variabile con diversi punti di divisione o probabilità (con il sottocomando `EXPECTED`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Test delle successioni

Il test delle successioni verifica se l'ordine delle occorrenze di due valori di una variabile è casuale. Una successione è una sequenza di osservazioni simili. Un campione con troppe o troppo poche successioni indica che il campione non è casuale.

Esempi. Si supponga che a venti persone venga chiesto se comprerebbero un determinato prodotto. La casualità prevista per il campione viene messa fortemente in dubbio se tutte le venti persone sono dello stesso sesso. È possibile utilizzare il test delle successioni per determinare se il campione è stato definito in modo casuale.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili.

Dati. Le variabili devono essere numeriche. Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma.

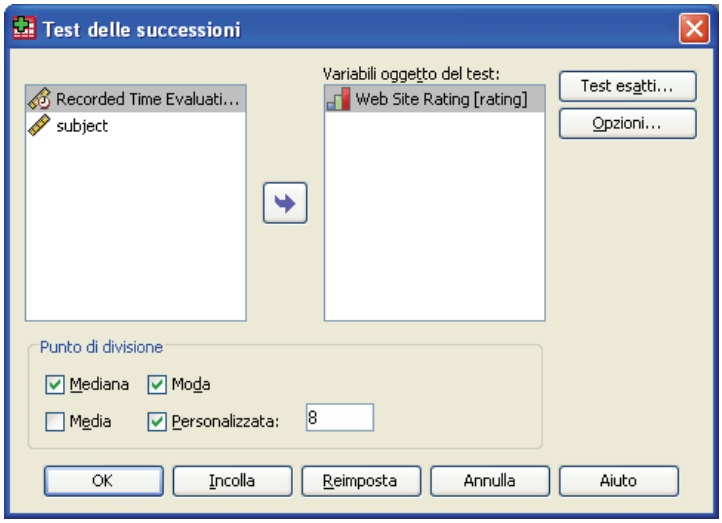
Assunzioni. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Utilizzare campioni da distribuzioni di probabilità continue.

Per ottenere un test delle successioni

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    Successioni...

Figura 39-5
Aggiunta di un punto di divisione personalizzato



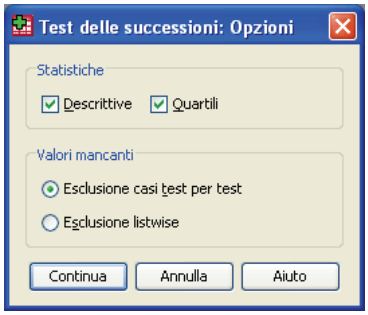
- Selezionare una o più variabili numeriche oggetto del test.
- È possibile fare clic su Opzioni per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Test delle successioni: Punto di divisione

Punto di divisione. Specifica un punto di divisione per dicotomizzare le variabili scelte dall'utente. È possibile utilizzare la media osservata, la mediana o la moda oppure un valore specificato come punto di divisione. I casi con valori minori del punto di divisione sono assegnati a un gruppo e i casi con valori uguali o maggiori del punto di divisione sono assegnati a un altro gruppo. Viene eseguito un test per ogni punto di divisione selezionato.

Test delle successioni: Opzioni

Figura 39-6
Finestra di dialogo Test delle successioni: Opzioni



Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (Test delle successioni)

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare diversi punti di divisione per diverse variabili (con il sottocomando `RUNS`).
- Verificare la stessa variabile rispetto a diversi punti di divisione personalizzati (con il sottocomando `RUNS`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Test di Kolmogorov-Smirnov per un campione

La procedura Test di Kolmogorov-Smirnov per un campione consente di confrontare la funzione di distribuzione cumulata osservata per una variabile con la distribuzione teorica specificata, che può essere normale, uniforme o di Poisson. La *Z* di Kolmogorov-Smirnov viene calcolata in base alla differenza maggiore (in valore assoluto) tra la funzione di distribuzione cumulata osservata e teorica. Questo test sulla bontà di adattamento permette di verificare se le osservazioni possono provenire dalla distribuzione specificata.

Esempio. Molti test parametrici richiedono variabili distribuite in modo normale. Il test di Kolmogorov-Smirnov per un campione può essere utilizzato per verificare che una variabile, ad esempio *reddito*, sia distribuita in modo normale.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili.

Dati. Utilizzare variabili quantitative (misurazione a livello di intervallo o di rapporto).

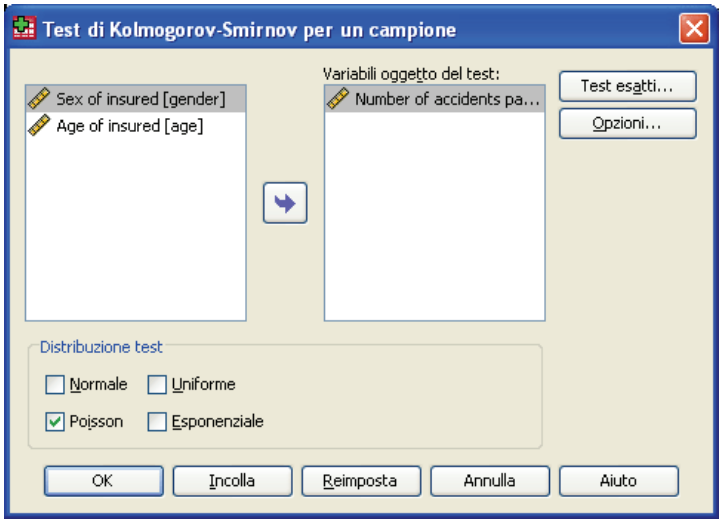
Assunzioni. Il test di Kolmogorov-Smirnov presume che i parametri della distribuzione del test vengano specificati anticipatamente. Questa procedura consente di valutare i parametri del campione. La media e la deviazione standard del campione sono i parametri della distribuzione normale, i valori minimo e massimo del campione definiscono l'intervallo di distribuzione uniforme e la media del campione è il parametro per la distribuzione Poisson e per la distribuzione esponenziale. La capacità del test di rilevare gli scostamenti dalla distribuzione ipotizzata può essere seriamente compromessa. Per effettuare test su una distribuzione normale con parametri stimati, è generalmente consigliabile usare il test K-S di Lilliefors corretto (selezionabile dalla procedura Esplora).

Per ottenere un test di Kolmogorov-Smirnov per un campione

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    K-S per 1 campione...

Figura 39-7
Finestra di dialogo Test di Kolmogorov-Smirnov per un campione

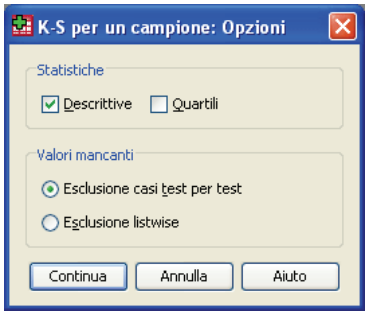


▶ Selezionare una o più variabili numeriche oggetto del test. Ogni variabile produce un test distinto.

▶ È possibile fare clic su Opzioni per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Test di Kolmogorov-Smirnov per un campione: Opzioni

Figura 39-8
Finestra di dialogo K-S per un campione



Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (Test di Kolmogorov-Smirnov per un campione)

Il linguaggio della sintassi dei comandi consente anche di specificare i parametri della distribuzione del test (con il sottocomando `K-S`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Test per due campioni indipendenti

La procedura del test per due campioni indipendenti consente di confrontare due gruppi di casi in base a una sola variabile.

Esempio. Sono stati creati nuovi apparecchi odontoiatrici che presentano numerosi vantaggi in termini di comodità, estetica ed efficacia ai fini dell'allineamento dei denti. Per determinare se i nuovi e i vecchi apparecchi devono essere portati per lo stesso periodo, sono stati scelti casualmente 10 bambini con il vecchio apparecchio e 10 bambini con il nuovo apparecchio. Dal test *U* di Mann-Whitney è possibile riscontrare che in media il nuovo apparecchio deve essere portato per un periodo di tempo inferiore.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: test *U* di Mann-Whitney, reazioni estreme di Moses, test *Z* di Kolmogorov-Smirnov, test delle successioni di Wald-Wolfowitz.

Dati. Utilizzare variabili numeriche che possono essere ordinate.

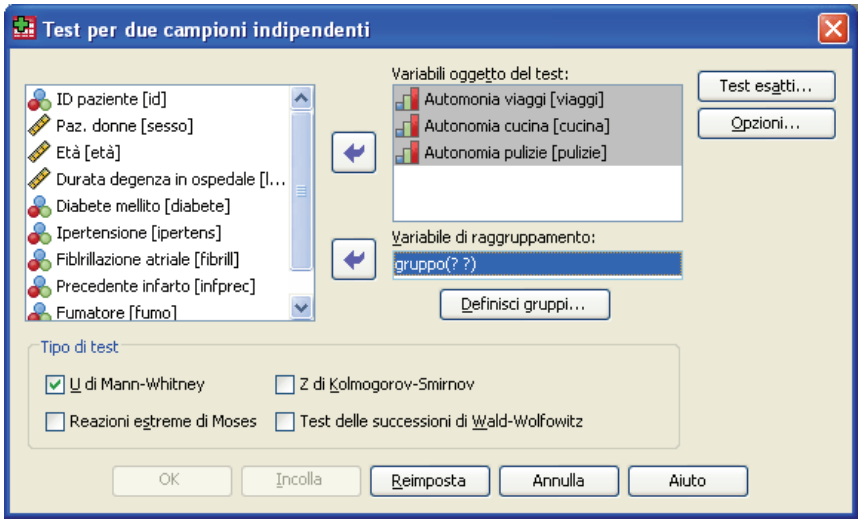
Assunzioni. Utilizzare campioni casuali indipendenti. Il test *U* di Mann-Whitney richiede l'utilizzo di due campioni di forma analoga.

Per ottenere un test per due campioni indipendenti

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    2 campioni indipendenti...

Figura 39-9
Finestra di dialogo Test per due campioni indipendenti



- Selezionare una o più variabili numeriche.
- Selezionare una variabile di raggruppamento e fare clic su Definisci gruppi per suddividere il file in due gruppi o campioni.

Tipi di test per due campioni indipendenti

Tipo di test. Sono disponibili quattro test che consentono di verificare se due campioni (gruppi) indipendenti provengono dalla stessa popolazione.

Il **test U di Mann-Whitney** è il test per due campioni indipendenti più diffuso. Equivale al test di Wilcoxon e al test di Kruskal-Wallis per due gruppi. Il test di Mann-Whitney permette di verificare l'equivalenza della posizione delle due popolazioni campione. Le osservazioni di entrambi i gruppi vengono combinate e ordinate per ranghi, assegnando la media dei ranghi ai valori a pari merito. Il numero di valori a pari merito deve essere inferiore al numero totale di osservazioni. Se la posizione delle popolazioni risulta identica, è necessario distribuire casualmente i ranghi tra i due campioni. Il test calcola il numero di volte in cui il punteggio del gruppo 1 è inferiore a quello del gruppo 2 e il numero di volte che un punteggio del gruppo 2 è inferiore al punteggio del gruppo 1. Il dato statistico che si ottiene con il test *U* di Mann-Whitney è inferiore a questi due numeri. Verrà visualizzata anche la statistica risultante dal test *W* di Wilcoxon, corrispondente alla somma dei ranghi del campione più piccolo. Se per entrambi i campioni è stato eseguito lo stesso numero di osservazioni, *W* è la somma dei ranghi del gruppo elencato per primo nella finestra di dialogo Due campioni indipendenti: Definisci gruppi.

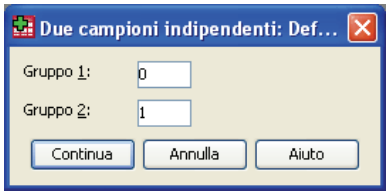
Il **test Z di Kolmogorov-Smirnov** e il **test delle successioni di Wald-Wolfowitz** sono test di carattere più generale che consentono di individuare le differenze tra le distribuzioni in termini di forma e posizione. Il test di Kolmogorov-Smirnov si basa sulla massima differenza in valore assoluto tra le funzioni di distribuzione cumulative osservate per entrambi i campioni. Quando tale differenza è significativa, le due distribuzioni vengono considerate diverse. Il test delle successioni di Wald-Wolfowitz consente di combinare e ordinare in ranghi le osservazioni di

entrambi i gruppi. Se i due campioni provengono dalla stessa popolazione, è necessario distribuire casualmente i gruppi all'interno della classifica.

Il **test delle reazioni estreme di Moses** si basa sull'ipotesi che la variabile sperimentale influenzi alcuni soggetti in una direzione e altri nella direzione opposta. Consente di verificare le risposte estreme confrontandole con un gruppo di controllo. Questo test è incentrato sull'estensione del gruppo di controllo e definisce la misura in cui i valori estremi del gruppo sperimentale influenzano l'estensione in caso di combinazione con il gruppo di controllo. Il gruppo di controllo viene definito dal valore del gruppo 1 nella finestra di dialogo Due campioni indipendenti: Definisci gruppi. Le osservazioni eseguite su entrambi i gruppi vengono combinate e classificate per ranghi. L'estensione del gruppo di controllo viene calcolata come la differenza tra i ranghi dei valori massimo e minimo del gruppo di controllo più 1. Poiché a causa di valori casuali anomali è probabile che l'intervallo dell'estensione risulti distorto, da ciascun estremo viene eliminato il 5% dei casi di controllo.

Test per due campioni indipendenti: Definisci gruppi

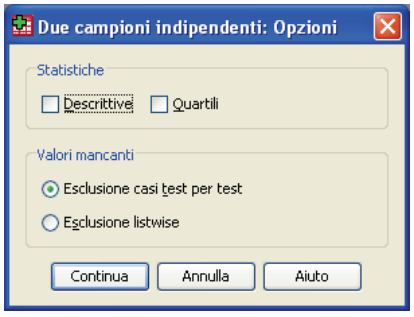
Figura 39-10
Finestra di dialogo Due campioni indipendenti: Definisci gruppi



Per suddividere il file in due gruppi o campioni, inserire un valore intero per il gruppo 1 e un altro per il gruppo 2. I casi con altri valori verranno esclusi dall'analisi.

Test per due campioni indipendenti: Opzioni

Figura 39-11
Finestra di dialogo Due campioni indipendenti: Opzioni



Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (Due campioni indipendenti)

Il linguaggio della sintassi dei comandi permette anche di specificare il numero di casi da eliminare dal test di Moses (con il sottocomando `MOSES`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Test per due campioni dipendenti

La procedura del test per due campioni dipendenti consente di confrontare la distribuzione di due variabili.

Esempio. In generale, le famiglie ricevono l'intero prezzo di offerta per la vendita della propria casa? Applicando il test di Wilcoxon ai dati relativi a 10 case, si riscontrerà che sette famiglie ricevono una somma inferiore al prezzo di offerta, una famiglia riceve una somma superiore, mentre due sole famiglie lo ricevono interamente.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: di Wilcoxon, del segno e di McNemar. Se è installata l'opzione Test esatti (disponibile solo nei sistemi operativi Windows), è disponibile anche il test di omogeneità marginale.

Dati. Utilizzare variabili numeriche che possono essere ordinate.

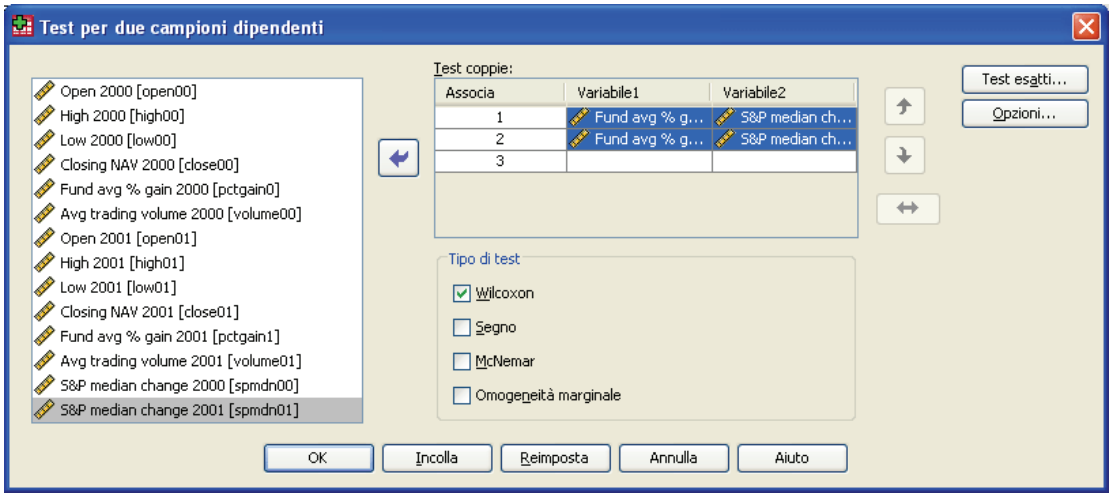
Assunzioni. Anche se per le due variabili non si ipotizza una particolare distribuzione, si presume che la distribuzione delle differenze a coppie sia simmetrica.

Per ottenere test per due campioni dipendenti

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    2 campioni dipendenti...

Figura 39-12
Finestra di dialogo Test per due campioni dipendenti



- Selezionare una o più coppie di variabili.

Tipi di test per due campioni dipendenti

I test descritti in questa sezione permettono di confrontare le distribuzioni di due variabili correlate. Il test più appropriato varia a seconda dei tipi di dati.

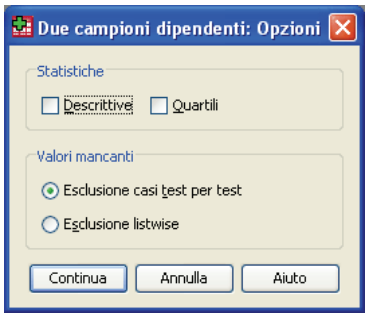
Se i dati sono continui, utilizzare il test del segno o di Wilcoxon. Il **test del segno** permette di calcolare le differenze tra le due variabili per tutti i casi e di classificarle come positive, negative o a pari merito. Se le due variabili sono distribuite in modo analogo, il numero di differenze positive e negative non differirà in misura significativa. Il **test di Wilcoxon** prende in considerazione le informazioni relative al segno e all'entità delle differenze tra le coppie. Poiché il test di Wilcoxon include un maggior numero di informazioni relative ai dati, risulta più valido del test del segno.

Se i dati sono binari, utilizzare il **test di McNemar**. Questo test viene in genere utilizzato in presenza di misure ripetute, ovvero quando la risposta del soggetto viene richiesta due volte: prima e dopo il verificarsi di un determinato evento. Il test di McNemar consente di determinare se il tasso di risposta iniziale (prima dell'evento) equivale al tasso di risposta finale (dopo l'evento). Questo test risulta particolarmente utile per individuare le variazioni della risposta in disegni sperimentali del tipo 'prima e dopo'.

Se i dati sono categoriali, utilizzare il **test di omogeneità marginale**. Estensione del test di McNemar dalla risposta binaria a quella multinomiale. Consente di verificare le variazioni della risposta utilizzando la distribuzione del chi-quadrato e risulta utile in disegni sperimentali del tipo 'prima e dopo'. Il test di omogeneità marginale è disponibile solo se è stato installato il modulo Exact Tests.

Test per due campioni dipendenti: Opzioni

Figura 39-13
Finestra di dialogo Due campioni dipendenti: Opzioni



Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (due campioni dipendenti)

Il linguaggio della sintassi dei comandi permette anche di verificare una variabile con ciascuna variabile dell'elenco.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Test per diversi campioni indipendenti

La procedura per i test per diversi campioni indipendenti consente di confrontare due o più gruppi di casi in base a una variabile.

Esempio. Le lampadine da 100 watt di tre diversi produttori si differenziano in relazione al tempo medio di bruciatura del filamento? Grazie all'ANOVA univariata di Kruskal-Wallis è possibile verificare che la durata media delle tre lampadine è effettivamente diversa.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: *H* di Kruskal-Wallis, mediana.

Dati. Utilizzare variabili numeriche che possono essere ordinate.

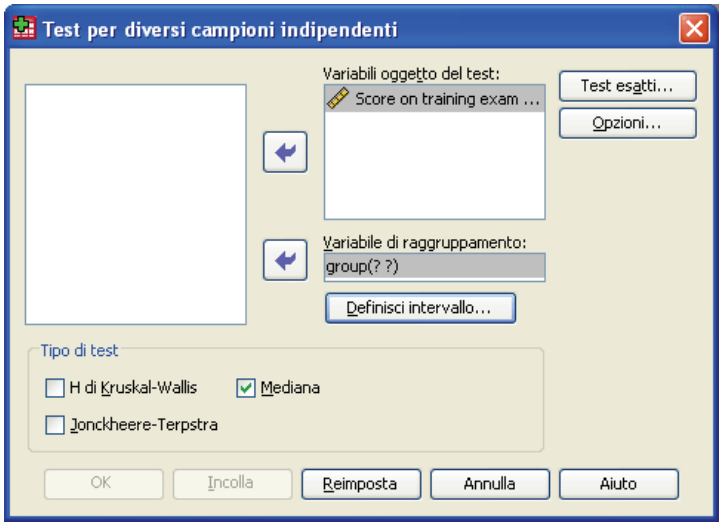
Assunzioni. Utilizzare campioni casuali indipendenti. IL test *H* di Kruskal-Wallis richiede che i campioni sottoposti a test siano simili per forma.

Per ottenere test per diversi campioni indipendenti

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    K campioni indipendenti...

Figura 39-14
Definizione del test della mediana



▶ Selezionare una o più variabili numeriche.

▶ Selezionare una variabile di raggruppamento e fare clic su Definisci intervallo per specificare i valori interi minimo e massimo per la variabile di raggruppamento.

Test per diversi campioni indipendenti: tipi di test

Sono disponibili tre test per stabilire se diversi campioni indipendenti sono stati estratti dalla stessa popolazione. Il test *H* di Kruskal-Wallis, il test della mediana e il test di Jonckheere-Terpstra consentono di verificare se i diversi campioni indipendenti sono stati estratti dalla stessa popolazione.

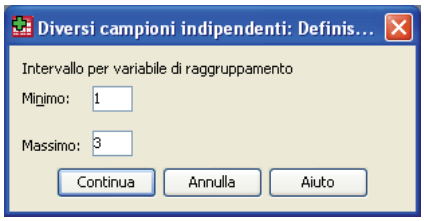
Il test **H di Kruskal-Wallis**, un'estensione del test *U* di Mann-Whitney, è la versione non parametrica dell'analisi univariata della varianza e consente di rilevare le differenze nella posizione di distribuzione. Il **test della mediana**, che è più generale ma non altrettanto potente, consente di rilevare le differenze distribuzionali nella posizione e nella forma. Il test *H* di Kruskal-Wallis e il test della mediana presumono che non esistano ordinamenti *a priori* delle *k* popolazioni da cui sono estratti i campioni.

Quando *esiste* un naturale ordinamento *a priori* (crescente o decrescente) delle *k* popolazioni, il **test di Jonckheere-Terpstra** è più potente. Ad esempio, le *k* popolazioni possono rappresentare *k* temperature crescenti. L'ipotesi che diverse temperature producano la stessa distribuzione della risposta è verificata rispetto all'ipotesi alternativa in base a cui al salire della temperatura, cresce il valore della risposta. Qui l'ipotesi alternativa è ordinata e quindi il test di Jonckheere-Terpstra è il

più appropriato da utilizzare. Il test di Jonckheere-Terpstra è disponibile solo se è installato il modulo aggiuntivo Testi esatti.

Test per diversi campioni indipendenti: Definisci intervallo

Figura 39-15
Finestra di dialogo Diversi campioni indipendenti: Definisci intervallo



Per definire l'intervallo, immettere valori interni per il minimo e il massimo che corrispondono alle categorie minore e maggiore della variabile di raggruppamento. Sono esclusi i casi con valori al di fuori dei limiti. Se, ad esempio, si specifica un limite inferiore di 1 e un limite superiore di 3, verranno utilizzati solo i valori interi compresi tra 1 e 3. Il valore minimo deve essere inferiore al valore massimo ed entrambi i valori devono essere specificati.

Test per diversi campioni indipendenti: Opzioni

Figura 39-16
Finestra di dialogo Diversi campioni indipendenti: Opzioni

Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (K campioni indipendenti)

Il linguaggio della sintassi dei comandi permette anche di specificare un valore diverso dalla mediana osservata per il test della mediana (con il sottocomando `MEDIAN`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Test per diversi campioni dipendenti

Il test per diversi campioni dipendenti consente di confrontare le distribuzioni di due o più variabili.

Esempio. Il pubblico associa diversi livello di prestigio al ruolo di dottore, avvocato, ufficiale della polizia e insegnante? A dieci persone viene chiesto di ordinare queste quattro occupazioni in base al prestigio. Il test di Friedman indica che il pubblico associa effettivamente livelli di prestigio diversi a queste quattro professioni.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: Friedman, *W* di Kendall e *Q* di Cochran.

Dati. Utilizzare variabili numeriche che possono essere ordinate.

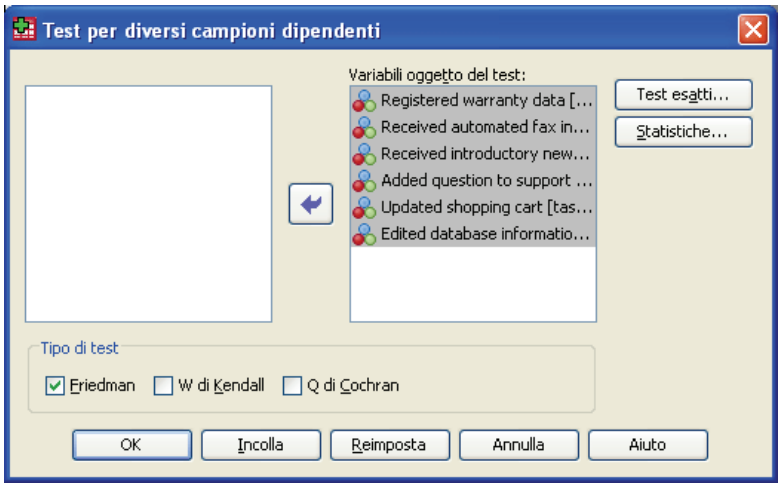
Assunzioni. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Utilizzare campioni casuali dipendenti.

Per ottenere i test per diversi campioni dipendenti

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    K campioni dipendenti...

Figura 39-17
Selezione di Cochran come tipo di test



▶ Selezionare una o più variabili oggetto del test numeriche.

Test per diversi campioni dipendenti: tipi di test

Sono disponibili tre test per confrontare le distribuzioni di diverse variabili correlate.

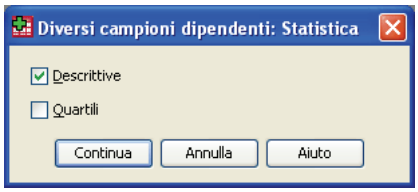
Il **testdi Friedman** è l'equivalente non parametrico di un disegno di misure ripetute per un campione o ANOVA a due vie con una osservazione per cella. Friedman verifica l'ipotesi nulla secondo cui *k* variabili correlate provengono dalla stessa popolazione. Per ogni caso, alle *k* variabili viene assegnato un rango da 1 a *k*. Le statistiche del test sono basate su questi ranghi.

W di Kendall è una normalizzazione delle statistiche di Friedman. È possibile interpretare il *W* di Kendall come il coefficiente di concordanza, che rappresenta la misura dell'accordo tra stimatori. Ogni caso è uno stimatore e ogni variabile è un elemento o individuo da stimare. Per ogni variabile viene calcolata la somma dei ranghi. *W* di Kendall varia tra 0 (nessun accordo) e 1 (accordo completo).

Q di Cochran è identico al test di Friedman ma è applicabile quando tutte le risposte sono binarie. Questo test è un'estensione del test di McNemar alla situazione di *k*-campioni. I test *Q* di Cochran verificano l'ipotesi secondo cui diverse variabili dicotomiche hanno la stessa media. Le variabili sono misurate sullo stesso individuo o su individui collegati fra loro.

Test per diversi campioni dipendenti: Statistica

Figura 39-18
Finestra di dialogo Diversi campioni dipendenti: Statistica



È possibile scegliere le statistiche.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Funzioni aggiuntive del comando NPAR TESTS (K campioni dipendenti)

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
40

Analisi a risposta multipla

Sono disponibili due procedure per l'analisi di insiemi a dicotomie e a categorie multiple. La procedura Risposte multiple: Frequenze consente di visualizzare le tabelle di frequenza. La procedura Risposte multiple: Tavole di contingenza consente di visualizzare tavole di contingenza a due o a tre dimensioni. Prima di utilizzare una delle procedure descritte, è necessario definire gli insiemi a risposta multipla.

Esempio. In questo esempio viene illustrato l'utilizzo degli elementi a risposta multipla in un'indagine di mercato. I dati sono fittizi e non devono essere interpretati come reali. È possibile condurre un'indagine tra i passeggeri di una linea aerea in volo su una particolare rotta per ottenere una valutazione della concorrenza. In questo esempio la compagnia American Airlines conduce un'indagine volta a rilevare se i propri passeggeri viaggiano con altre linee aeree sulla rotta Chicago-New York e a determinare l'importanza relativa dei fattori di programmazione e di servizio ai fini della scelta della linea aerea. Al momento dell'imbarco, l'assistente di volo consegna a ciascun passeggero un breve questionario. La prima domanda è la seguente. Tra le seguenti compagnie aeree, contrassegnare quelle utilizzate su questa stessa rotta almeno una volta negli ultimi sei mesi: American, United, TWA, USAir, Altro. Si tratta di una domanda a risposta multipla in quanto il passeggero può indicare più risposte. La domanda, tuttavia, non può essere codificata direttamente in quanto una variabile può contenere un solo valore per ciascun caso. È necessario utilizzare più variabili per associare le risposte a ciascuna domanda. Per eseguire questa operazione è possibile procedere in due modi. Il primo modo consiste nel definire una variabile per ciascuna delle scelte (ad esempio, American, United, TWA, USAir e Altro). Se il passeggero indica la linea United, alla variabile *united* verrà assegnato il codice 1 e in caso contrario il codice 0. Si tratta di un **metodo a dicotomie multiple** per l'assegnazione delle variabili. Il secondo metodo per la classificazione delle risposte è il **metodo a categorie multiple**, che consente di valutare il numero massimo di risposte possibili alla domanda e di impostare lo stesso numero di variabili, con i codici utilizzati per specificare la linea aerea utilizzata. Dall'esame di un campione di questionari può risultare che negli ultimi sei mesi nessun utente ha viaggiato con più di tre diverse linee aeree su questa rotta. Può inoltre risultare che, a causa della deregulation delle linee aree, nella categoria Altro ne vengano citate altre 10. Utilizzando il metodo a risposta multipla, vengono definite tre variabili, ciascuna delle quali è codificata come 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta* e così via. Se un determinato passeggero indica American e TWA, alla prima variabile viene assegnato il codice 1, alla seconda il codice 3 e alla terza un codice di valore mancante. Un altro passeggero può aver indicato American e specificato Delta. In questo caso, alla prima variabile viene assegnato il codice 1, alla seconda 5 e alla terza un codice di valore mancante. Se invece si utilizza il metodo a dicotomie multiple, si otterranno 14 variabili distinte. Sebbene ai fini di questa indagine sia possibile utilizzare entrambi i metodi, la scelta del metodo dipende dalla distribuzione delle risposte.

Risposte multiple: Definisci insiemi

La procedura di definizione degli insiemi di variabili a risposta multipla consente di raggruppare le variabili elementari in insiemi a dicotomie o a categorie multiple, per i quali è possibile ottenere tabelle di frequenza e tavole di contingenza. È possibile definire fino a 20 insiemi a risposta multipla. A ogni insieme è necessario assegnare un nome univoco. Per rimuovere un insieme, evidenziarlo nell'elenco dei gruppi a risposta multipla e quindi scegliere Rimuovi. Per modificare un insieme, evidenziarlo nell'elenco, modificare le caratteristiche di definizione desiderate e quindi scegliere Cambia.

È possibile codificare le variabili elementari come dicotomie o categorie. Per l'utilizzo delle variabili dicotomiche, selezionare Dicotomie per creare un insieme a dicotomie multiple. Specificare un valore intero nella casella Valore conteggiato. Ciascuna variabile contenente almeno un'occorrenza del valore conteggiato diventa una categoria dell'insieme a dicotomie multiple. Selezionare Categorie per creare un insieme a categorie multiple con lo stesso intervallo di valori delle variabili che lo compongono. Specificare valori interi come valori minimo e massimo dell'intervallo di categorie dell'insieme a categorie multiple. Verrà calcolato il totale di ogni singolo valore intero nell'intervallo per tutte le variabili. Le categorie vuote non verranno incluse nella tabella.

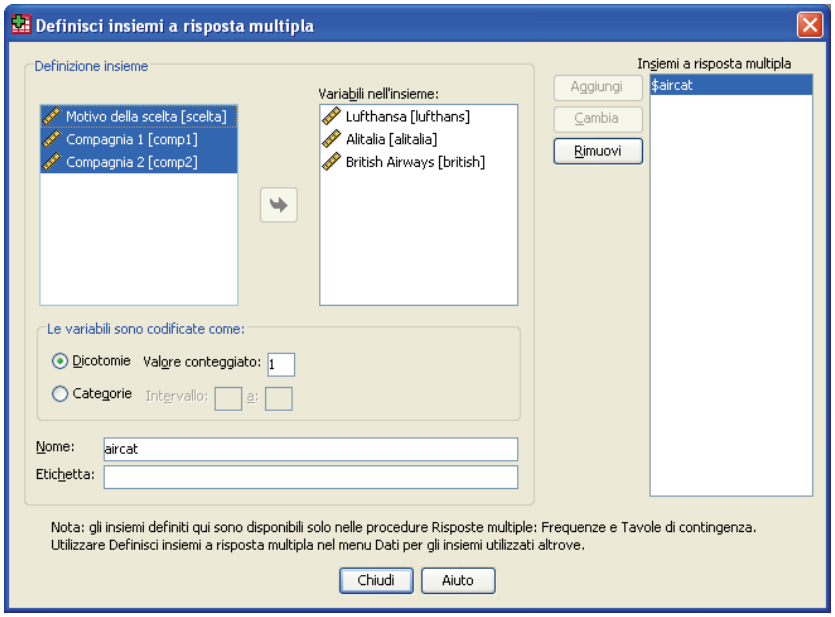
A ogni insieme a risposta multipla deve essere assegnato un nome univoco composto al massimo da sette caratteri. Al nome assegnato verrà aggiunto automaticamente il prefisso $ (segno di dollaro). Non è possibile utilizzare i seguenti nomi riservati: *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* e *width*. Il nome dell'insieme a risposta multipla esiste solo ai fini dell'utilizzo in procedure a risposta multipla. Non è possibile fare riferimento ai nomi di insiemi a risposta multipla in altre procedure. È inoltre possibile inserire un'etichetta di variabile descrittiva per l'insieme a risposta multipla. L'etichetta può essere costituita al massimo da 40 caratteri.

Per definire gli insiemi a risposta multipla

▶ Dai menu, scegliere:

Analizza
  Risposte multiple
    Definisci insiemi...

Figura 40-1
Finestra di dialogo Insiemi a risposta multipla



- Selezionare due o più variabili.
- Se le variabili sono codificate come dicotomie, indicare il valore che si desidera calcolare. Se le variabili sono codificate come categorie, definire l'intervallo delle categorie.
- Immettere un nome univoco per ciascun insieme a risposta multipla.
- Scegliere Aggiungi per aggiungere l'insieme a risposta multipla all'elenco di insiemi definiti.

Risposte multiple: Frequenze

La procedura Risposte multiple: Frequenze consente di ottenere tabelle di frequenza per gli insiemi a risposta multipla. È innanzitutto necessario definire uno o più insiemi a risposta multipla (vedere "Risposte multiple: Definisci insiemi").

Per gli insiemi a dicotomie multiple, i nomi delle categorie indicati nell'output vengono determinati in base alle etichette definite per le variabili elementari del gruppo. Se le etichette delle variabili non sono definite, i nomi delle variabili verranno utilizzati come etichette. Per gli insiemi a categorie multiple, le etichette di categoria vengono determinate in base alle etichette dei valori della prima variabile del gruppo. Se le categorie mancanti per la prima variabile sono presenti per altre variabili del gruppo, definire un'etichetta dei valori per le categorie mancanti.

Valori mancanti. I casi con valori mancanti vengono esclusi tabella per tabella. È inoltre possibile scegliere una delle seguenti opzioni o entrambe:

- **Esclusione listwise all'interno delle dicotomie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabella dell'insieme a dicotomie multiple. Questa opzione può essere applicata solo agli insiemi a risposta multipla definiti come insiemi dicotomici. Per impostazione predefinita, un caso viene considerato mancante per un insieme a dicotomie multiple se nessuna delle variabili che lo compongono contiene il valore conteggiato. I casi

con valori mancanti solo per alcune variabili verranno inclusi nelle tabelle del gruppo se almeno una variabile contiene il valore conteggiato.

- **Esclusione listwise all'interno delle categorie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabella dell'insieme a categorie multiple. Questa opzione viene applicata solo a insiemi a risposta multipla definiti come insiemi di categorie. Per impostazione predefinita, un caso viene considerato mancante per un insieme a categorie multiple solo se nessuno dei componenti contiene valori validi all'interno dell'intervallo definito.

Esempio. Ogni variabile creata in base a una domanda di un questionario è una variabile elementare. Per analizzare un elemento di un insieme a risposta multipla, è necessario unire le variabili in uno dei due tipi di insiemi a risposta multipla: un insieme a dicotomie multiple oppure un insieme a categorie multiple. Se ad esempio in un'indagine sulle linee aeree dove viene richiesto con quale delle tre linee aeree indicate (American, United, TWA) si è viaggiato negli ultimi sei mesi sono state utilizzate variabili dicotomiche e si è definito un **insieme a dicotomie multiple**, ciascuna delle tre variabili dell'insieme diventerà una categoria della variabile di gruppo. I conteggi e le percentuali relativi alle tre linee aeree verranno visualizzati in una sola tabella di frequenza. Se risulta che nessuna persona ha indicato più di due linee aeree, è possibile creare due variabili, ciascuna con tre codici, ovvero uno per ogni linea aerea. Se si definisce un **insieme a categorie multiple**, i valori verranno inseriti nella tabella aggiungendo gli stessi codici alle variabili elementari. L'insieme di valori risultante equivale agli insiemi di ciascuna variabile elementare. Trenta risposte per United rappresentano ad esempio la somma delle cinque risposte per United per la linea aerea 1 e delle venticinque risposte per United per la linea aerea 2. I conteggi e le percentuali relativi alle tre linee aeree verranno visualizzati in una sola tabella di frequenza.

Statistiche. Tabelle di frequenza in cui vengono visualizzati i conteggi, le percentuali delle risposte, le percentuali dei casi, il numero di casi validi e il numero di casi mancanti.

Dati. Utilizzare gli insiemi a risposta multipla.

Assunzioni. I conteggi e le percentuali forniscono un'utile descrizione dei dati provenienti da qualsiasi distribuzione.

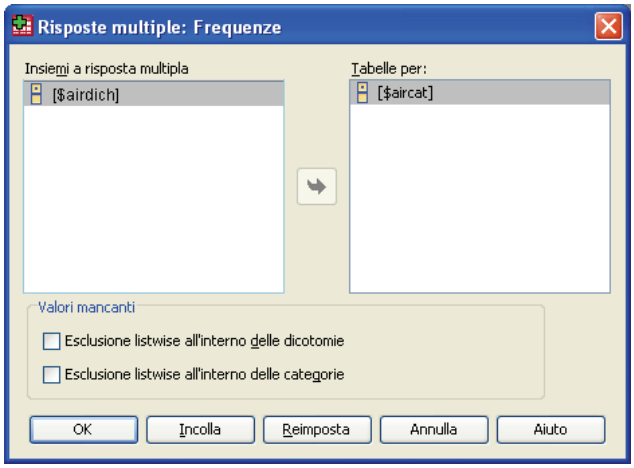
Procedure correlate. La procedura Risposte multiple: Definisci insiemi consente di definire insiemi a risposta multipla.

Per ottenere le frequenze delle risposte multiple

▶ Dai menu, scegliere:

Analizza
  Risposte multiple
    Frequenze...

Figura 40-2
Finestra di dialogo Risposte multiple: Frequenze



▶ Selezionare uno o più insiemi a risposta multipla.

Risposte multiple: Tavole di contingenza

La procedura Risposte multiple: Tavole di contingenza consente di incrociare insiemi a risposta multipla definiti, variabili elementari o una combinazione di entrambi. È inoltre possibile ottenere le percentuali nelle celle in base a casi o risposte, modificare il trattamento dei valori mancanti oppure accoppiare le variabili di insiemi diversi. È innanzitutto necessario definire uno o più insiemi a risposta multipla (vedere "Per definire gli insiemi a risposta multipla").

Per gli insiemi a dicotomie multiple, i nomi delle categorie indicati nell'output vengono determinati in base alle etichette definite per le variabili elementari del gruppo. Se le etichette delle variabili non sono definite, i nomi delle variabili verranno utilizzati come etichette. Per gli insiemi a categorie multiple, le etichette di categoria vengono determinate in base alle etichette dei valori della prima variabile del gruppo. Se le categorie mancanti per la prima variabile sono presenti per altre variabili del gruppo, definire un'etichetta dei valori per le categorie mancanti. Le etichette delle categorie per le colonne verranno visualizzate su tre righe, composte al massimo da otto caratteri ciascuna. Per evitare di troncare le parole, è possibile invertire le righe e le colonne oppure ridefinire le etichette.

Esempio. In questa procedura è possibile incrociare gli insiemi a dicotomie e a categorie multiple con altre variabili. In un questionario riservato ai passeggeri delle linee aeree venivano richieste le seguenti informazioni: Contrassegnare tra le seguenti tutte le linee aeree di cui ci si è serviti almeno una volta negli ultimi sei mesi (American, United, TWA). Cosa è più importante per la scelta di un volo, la programmazione o il servizio offerto? Scegliere una sola opzione. Dopo aver inserito i dati come dicotomie o categorie multiple e averli uniti in un insieme, è possibile incrociare le domande relative alle linee aeree con quelle relative al servizio o alla programmazione.

Statistiche. Tavole di contingenza con conteggi relativi a celle, righe, colonne e totale e percentuali relative a celle, righe, colonne e totale. Le percentuali nelle celle possono basarsi sui casi o sulle risposte.

Dati. Utilizzare insiemi a risposta multipla oppure variabili categoriali numeriche.

Assunzioni. I conteggi e le percentuali forniscono un'utile descrizione dei dati provenienti da qualsiasi distribuzione.

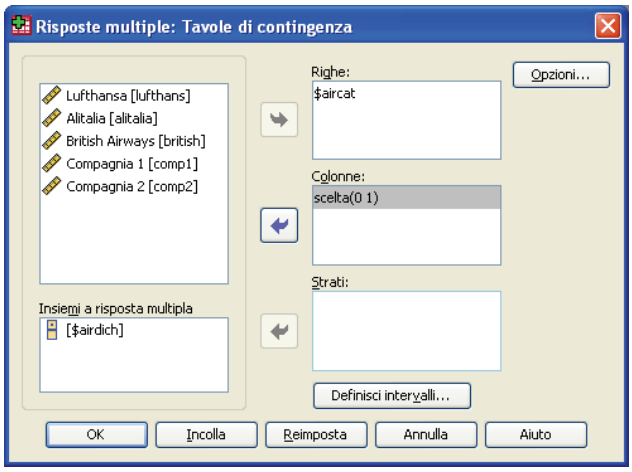
Procedure correlate. La procedura Risposte multiple: Definisci insiemi consente di definire insiemi a risposta multipla.

Per ottenere tavole di contingenza a risposta multipla

▶ Dai menu, scegliere:

Analizza
  Risposte multiple
    Tavole di contingenza...

Figura 40-3
Finestra di dialogo Risposte multiple: Tavole di contingenza



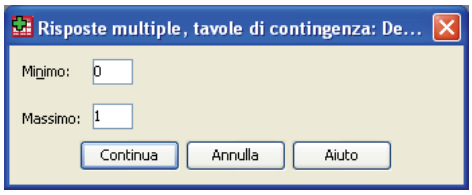
▶ Selezionare una o più variabili numeriche o insiemi a risposta multipla per ciascuna dimensione delle tavole di contingenza.

▶ Definire l'intervallo di ciascuna variabile elementare.

È inoltre possibile ottenere una tavola di contingenza a due vie per ciascuna categoria di una variabile di controllo o di un insieme a risposta multipla. Selezionare uno o più elementi dall'elenco Strati.

Risposte multiple, tavole di contingenza: Definisci intervalli delle variabili

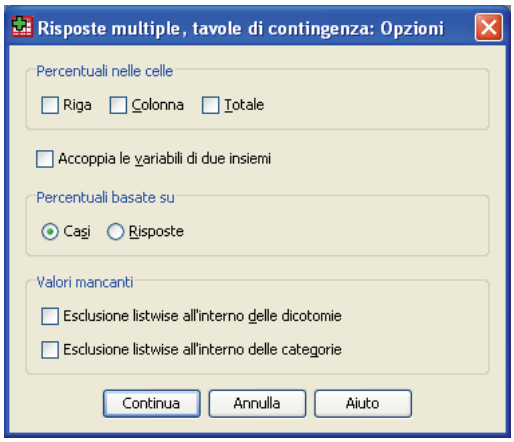
Figura 40-4
Finestra di dialogo Risposte multiple, tavole di contingenza: Definisci intervalli delle variabili



È necessario definire gli intervalli dei valori per tutte le variabili elementari delle tavole di contingenza. Specificare il valore di categoria minimo e massimo che si desidera inserire nelle tavole di contingenza. Le categorie che non rientrano nell'intervallo verranno escluse dall'analisi. Si assume che i valori inclusi nell'intervallo siano interi (i non interi verranno troncati).

Risposte multiple, tavole di contingenza: Opzioni

Figura 40-5
Finestra di dialogo Risposte multiple, tavole di contingenza: Opzioni



Percentuali nelle celle. I conteggi di cella vengono sempre visualizzati. È possibile impostare la visualizzazione di percentuali di riga, percentuali di colonna e percentuali per tabelle a due vie (totale).

Percentuali basate su. È possibile basare le percentuali nelle celle sui casi (o persone che rispondono). Questa opzione non è disponibile se si seleziona la corrispondenza delle variabili tra insiemi a categorie multiple. È inoltre possibile basare le percentuali nelle celle sulle risposte. Per gli insiemi a dicotomie multiple, il numero di risposte equivale al numero di valori conteggiati nei diversi casi. Per gli insiemi a categorie multiple, il numero di risposte equivale al numero di valori dell'intervallo definito.

Valori mancanti. È possibile scegliere una delle seguenti opzioni o entrambe:

- **Esclusione listwise all'interno delle dicotomie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabella dell'insieme a dicotomie multiple. Questa opzione può essere applicata solo agli insiemi a risposta multipla definiti come insiemi dicotomici. Per impostazione predefinita, un caso viene considerato mancante per un insieme a dicotomie multiple se nessuna delle variabili che lo compongono contiene il valore conteggiato. I casi con valori mancanti solo per alcune variabili verranno inclusi nelle tabelle del gruppo se almeno una variabile contiene il valore conteggiato.
- **Esclusione listwise all'interno delle categorie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabella dell'insieme a categorie multiple. Questa opzione viene applicata solo a insiemi a risposta multipla definiti come insiemi di categorie. Per impostazione predefinita, un caso viene considerato mancante per un insieme a categorie multiple solo se nessuno dei componenti contiene valori validi all'interno dell'intervallo definito.

Per impostazione predefinita, quando si incrociano due insiemi a categorie multiple, ciascuna variabile del primo gruppo verrà incrociata con ciascuna variabile del secondo gruppo e quindi verranno sommati i conteggi relativi a ciascuna cella. Alcune risposte, pertanto, potranno comparire più volte nella stessa tabella. È possibile scegliere la seguente opzione:

Accoppia le variabili di due insiemi. Consente di associare la prima variabile del primo gruppo con la prima variabile del secondo gruppo e così via. Se viene selezionata questa opzione, le percentuali nelle celle saranno basate sulle risposte e non sulle persone che rispondono. Questa opzione non è disponibile per gli insiemi a dicotomie multiple né per le variabili elementari.

Funzioni aggiuntive del comando MULT RESPONSE

Il linguaggio della sintassi dei comandi consente inoltre di:

- Ottenere tavole di contingenza con un massimo di cinque dimensioni (con il sottocomando `BY`).
- Modificare le opzioni di formattazione dell'output, inclusa l'eliminazione delle etichette dei valori (con il sottocomando `FORMAT`).

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
41

Risultati di report

Gli elenchi dei casi e le statistiche descrittive sono strumenti fondamentali per lo studio e la presentazione dei dati. Per creare elenchi dei casi è possibile utilizzare l'Editor dei dati o la procedura Riassumi, per produrre conteggi di frequenze e statistiche descrittive è possibile utilizzare la procedura Frequenze, mentre per creare statistiche per la sottopopolazione è possibile utilizzare la procedura Medie. Queste procedure utilizzano un formato progettato per rendere chiare le informazioni. Per visualizzare le informazioni in un formato diverso, è possibile impostare la presentazione dei dati mediante le opzioni per i report Riepiloghi per righe e Riepiloghi per colonne.

Report: Riepiloghi per righe

La procedura Report: Riepiloghi per righe consente di creare report in cui statistiche riassuntive diverse sono disposte in righe distinte. Sono inoltre disponibili gli elenchi dei casi, che possono includere o meno le statistiche riassuntive.

Esempio. Una ditta proprietaria di una catena di negozi al dettaglio registra le informazioni sui dipendenti, che includono informazioni sugli stipendi e le mansioni nonché sul negozio e il reparto in cui lavora ogni dipendente. È quindi possibile creare un report che includa le informazioni relative a ogni impiegato (elenco) suddivise per negozio e per reparto (variabili di separazione) e che includa statistiche riassuntive (ad esempio, lo stipendio medio) per ciascun negozio e reparto nonché per ciascun reparto di ogni negozio.

Colonne dati. Elenca le variabili da rappresentare nel report, per le quali si desidera creare elenchi dei casi o statistiche riassuntive e controlla il formato per la visualizzazione delle colonne di dati.

Colonne di separazione. Elenca le variabili di separazione facoltative che suddividono il report in più gruppi e consente di gestire le statistiche riassuntive e i formati per la visualizzazione delle colonne di separazione. Se sono presenti più variabili di separazione, per ogni categoria di ciascuna variabile di separazione verrà creato un gruppo distinto all'interno delle categorie della variabile di separazione precedente nell'elenco. Le variabili di separazione devono essere variabili categoriali discrete che suddividono i casi in un numero limitato di categorie significative. I singoli valori di ciascuna variabile di separazione vengono visualizzati, ordinati, in una colonna distinta a sinistra di tutte le colonne dati.

Report. Consente di controllare le caratteristiche generali del report, inclusi i titoli, le statistiche riassuntive globali, la visualizzazione dei valori mancanti e la numerazione delle pagine.

Visualizza casi. Consente di visualizzare i valori effettivi (o etichette dei valori) delle variabili delle colonne di dati per ciascun caso. In tal modo viene creato un listato che potrebbe risultare molto più lungo di un report riepilogativo.

Bozza. Consente di visualizzare solo la prima pagina del report. Questa opzione è utile per visualizzare in anteprima il formato del report prima di generarlo.

I dati sono già ordinati. Se il report include variabili di separazione, prima di generarlo è necessario ordinare il file dati in base ai valori delle variabili di separazione. Se il file di dati è già ordinato in base ai valori delle variabili di separazione, è possibile ridurre i tempi di elaborazione selezionando questa opzione. Questa opzione risulta particolarmente utile dopo aver visualizzato un'anteprima del report.

Per ottenere un riepilogo: Riepiloghi per righe

- Dai menu, scegliere:
  Analizza
  Report
  Report: Riepiloghi per righe...

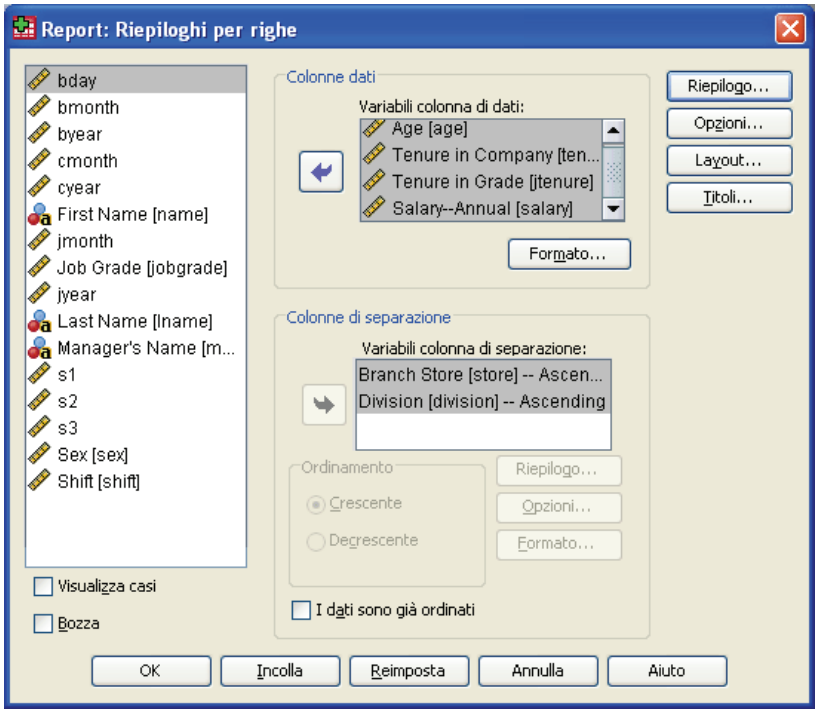
- Selezionare una o più variabili per Colonne dati. Per ogni variabile selezionata verrà generata una colonna nel report.

- Per i report ordinati e visualizzati in base ai sottogruppi, selezionare una o più variabili per Colonne di separazione.

- Per i report con statistiche riassuntive per i sottogruppi definiti in base alle variabili di separazione, selezionare la variabile di separazione nell'elenco Variabili colonna di separazione e fare clic su Riepilogo nel gruppo Colonne di separazione per specificare le misure riassuntive.

- Per i report con statistiche riassuntive globali, fare clic su Riepilogo per specificare le misure riassuntive.

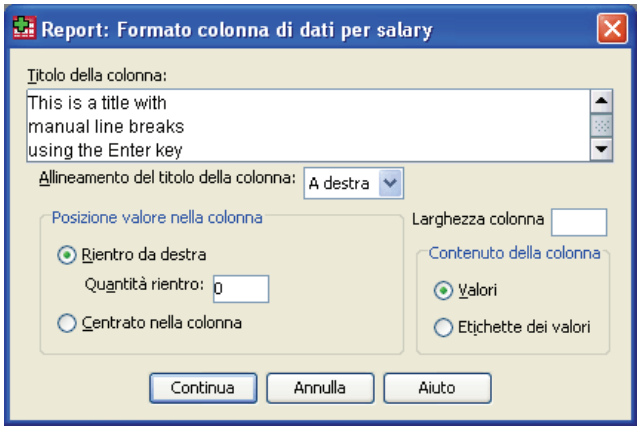
Figura 41-1
Finestra di dialogo Report: Riepiloghi per righe



Formato delle colonne e di separazione del report

Nelle finestre di dialogo relative al formato è possibile impostare i titoli e la larghezza delle colonne, l'allineamento del testo e la visualizzazione dei valori e delle etichette dei valori. L'opzione Formato colonna dati consente di impostare il formato delle colonne dati nella parte destra della pagina del report. L'opzione Formato di separazione consente di impostare il formato delle colonne di separazione nella parte sinistra.

Figura 41-2
Finestra di dialogo Report: Formato colonna dati



Titolo della colonna. Consente di impostare il titolo della colonna per la variabile selezionata. I titoli lunghi vanno a capo automaticamente all'interno della colonna. Per inserire manualmente le separazioni di riga nella posizione in cui si desidera che i titoli vadano a capo, è possibile utilizzare il tasto Invio.

Posizione valore nella colonna. Per la variabile selezionata, consente di impostare l'allineamento dei valori o delle etichette dei valori all'interno della colonna. L'allineamento dei valori o delle etichette non modifica l'allineamento delle intestazioni di colonna. È possibile rientrare il contenuto delle colonne di un numero specifico di caratteri oppure centrarlo.

Contenuto della colonna. Per la variabile selezionata, consente di impostare la visualizzazione dei valori o delle etichette dei valori definite. I valori per i quali non è stata definita alcuna etichetta vengono sempre visualizzati. Non è disponibile per le colonne dati nei report di riepilogo per colonne.

Report: Linee riassuntive per/Linee riassuntive finali

Le due finestre di dialogo Report: Linee riassuntive consentono di impostare la visualizzazione delle statistiche riassuntive per i gruppi di interruzione e per l'intero report. Linee riassuntive consente di impostare le statistiche dei sottogruppi per ciascuna categoria definita tramite le variabili di separazione. Linee riassuntive finali consente di impostare le statistiche globali visualizzate nella parte finale del report.

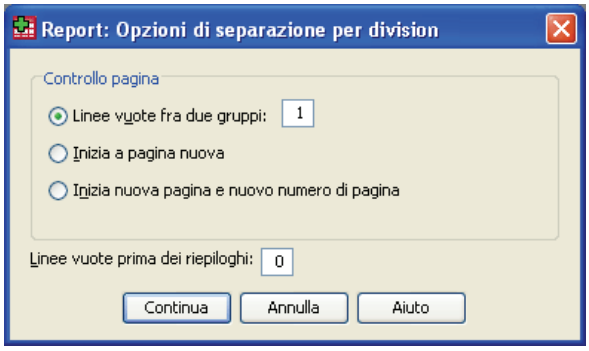
Figura 41-3
Finestra di dialogo Report: Linee riassuntive

Le statistiche riassuntive disponibili sono: somma, media, minimo, massimo, numero di casi, percentuale di casi al di sopra o al di sotto di un valore specifico, percentuale di casi entro un intervallo specifico di valori, deviazione standard, curtosi, varianza e asimmetria.

Report: Opzioni di separazione

La funzione Opzioni di separazione consente di impostare la spaziatura e l'impaginazione delle informazioni sui gruppi.

Figura 41-4
Finestra di dialogo Report: Opzioni di separazione



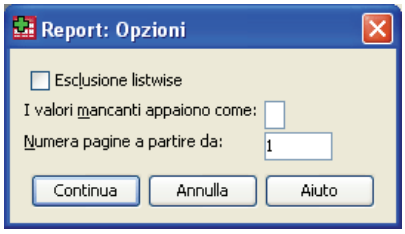
Controllo pagina. Consente di impostare la spaziatura e l'impaginazione per le categorie relative alla variabile di separazione selezionata. È possibile impostare il numero desiderato di righe vuote tra i gruppi o fare in modo che ciascun gruppo inizi in una nuova pagina.

Linee vuote prima dei riepiloghi. Consente di impostare il numero delle linee vuote tra le etichette dell'asse delle categorie o i dati e le statistiche riassuntive. Questa funzionalità risulta particolarmente utile per i report combinati che includono sia l'elenco dei singoli casi che le statistiche riassuntive per i gruppi. In questo tipo di report è possibile inserire una spaziatura tra l'elenco dei casi e le statistiche riassuntive.

Report: Opzioni

La funzione Report: Opzioni consente di impostare la modalità di elaborazione e la visualizzazione dei valori mancanti e la numerazione delle pagine del report.

Figura 41-5
Finestra di dialogo Report: Opzioni



Esclusione listwise. Consente di eliminare dal report i casi con valori mancanti per le variabili del report.

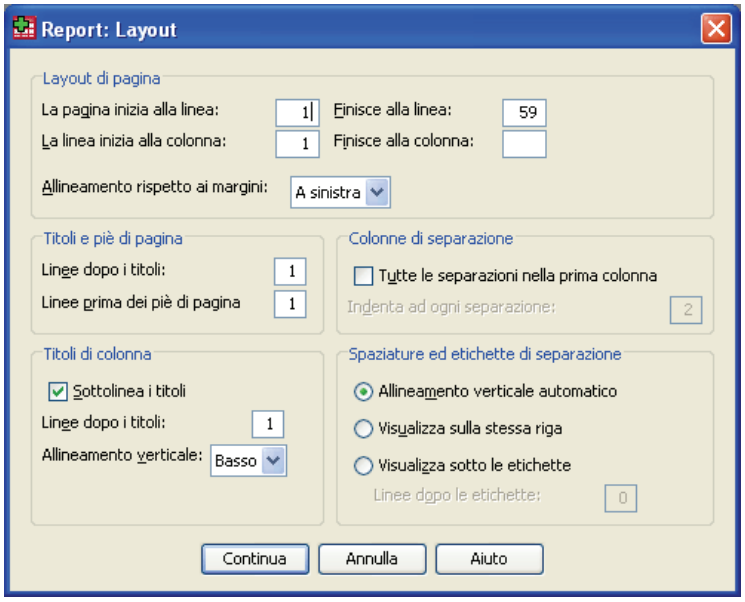
I valori mancanti appaiono come. Consente di specificare il simbolo che rappresenta i valori mancanti nel file dati. Il simbolo deve essere costituito da un solo carattere e viene usato per rappresentare sia i **valori mancanti di sistema** che i **valori mancanti definiti dall'utente**.

Numera pagine a partire da. Consente di specificare un numero di pagina con cui contrassegnare la prima pagina del report.

Report: Layout

L'opzione Report: Layout consente di impostare la larghezza e la lunghezza di ogni pagina del report, la posizione del report nella pagina e l'inserimento di linee vuote ed etichette.

Figura 41-6
Finestra di dialogo Report: Layout



Layout di pagina. Consente di impostare i margini della pagina espressi in linee (superiori e inferiori) e caratteri (a destra e a sinistra) nonché l'allineamento dei report all'interno dei margini.

Titoli e piè di pagina. Consente di impostare il numero delle linee che separano i piè di pagina e i titoli della pagina dal corpo del report.

Colonne di separazione. Consente di impostare la visualizzazione delle colonne di separazione. Se vengono specificate più variabili di separazione, queste possono trovarsi in colonne diverse oppure nella prima colonna. Se tutte le variabili di separazione vengono inserite nella prima colonna, verrà creato un report di larghezza minore.

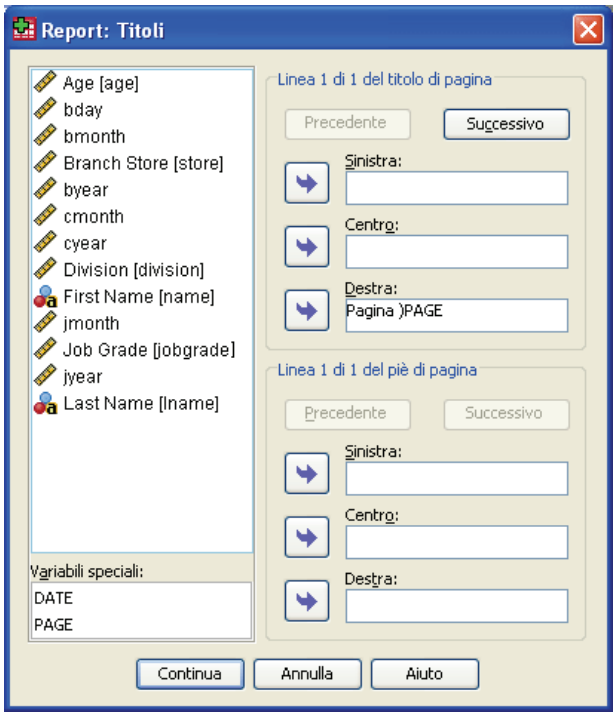
Titoli di colonna. Consente di impostare la visualizzazione dei titoli di colonna, inclusi lo spazio tra i titoli e il corpo del report, la sottolineatura del titolo e l'allineamento verticale dei titoli di colonna.

Spaziature ed etichette di separazione. Consente di posizionare le informazioni relative alle colonne dati (valori e/o statistiche riassuntive) in relazione alle etichette di separazione all'inizio di ogni gruppo. La prima riga delle informazioni sulla colonna dati può iniziare sulla stessa riga in cui si trova l'etichetta di gruppo o dopo un numero specifico di righe rispetto alla posizione dell'etichetta di gruppo. Non è disponibile per i report di riepilogo per colonne.

Report: Titoli

L'opzione Report: Titoli consente di impostare il contenuto e la posizione dei titoli e dei piè di pagina dei report. È possibile specificare fino a dieci righe per i titoli di pagina e per i piè di pagina, con componenti centrati oppure allineati a destra o a sinistra in ciascuna riga.

Figura 41-7
Finestra di dialogo Report: Titoli



Se si inseriscono variabili in titoli e in piè di pagina, l'etichetta dei valori corrente o il valore della variabile verrà visualizzato nel titolo o nel piè di pagina. Nei titoli viene visualizzata l'etichetta dei valori corrispondente al valore della variabile all'inizio della pagina. Nei piè di pagina viene visualizzata l'etichetta dei valori corrispondente al valore della variabile alla fine della pagina. Se non sono presenti etichette dei valori, viene visualizzato il valore effettivo.

Variabili speciali. Le variabili speciali *DATE* e *PAGE* consentono di inserire la data corrente o il numero di pagina in una delle righe dell'intestazione o del piè di pagina del report. Se il file dati utilizzato contiene le variabili denominate *DATE* o *PAGE*, non sarà possibile utilizzare tali variabili nei titoli o nei piè di pagina.

Report: Riepiloghi per colonne

L'opzione Report: Riepiloghi per colonne consente di creare report di riepilogo in cui le statistiche riassuntive vengono visualizzate in colonne distinte.

Esempio. Una ditta proprietaria di una catena di negozi al dettaglio registra le informazioni sui dipendenti. Tra queste sono comprese informazioni sugli stipendi e le mansioni nonché sul reparto in cui lavora ogni dipendente. È quindi possibile creare un report che includa le statistiche riassuntive sugli stipendi (ad esempio media, minimo e massimo) per ogni reparto.

Colonne dati. Consente di visualizzare un elenco delle variabili da rappresentare nel report e per le quali si desidera produrre statistiche riassuntive e di impostare il formato di visualizzazione e le statistiche riassuntive visualizzate per ogni variabile.

Colonne di separazione. Consente di visualizzare l'elenco delle variabili di separazione facoltative che suddividono il report in più gruppi e di impostare i formati di visualizzazione delle colonne di separazione. Se sono presenti più variabili di separazione, per ogni categoria di ciascuna variabile di separazione verrà creato un gruppo distinto all'interno delle categorie della variabile di separazione precedente nell'elenco. Le variabili di separazione devono essere variabili categoriali discrete che suddividono i casi in un numero limitato di categorie significative.

Report. Consente di impostare le caratteristiche globali del report, inclusi i titoli, la visualizzazione dei valori mancanti e la numerazione delle pagine.

Bozza. Consente di visualizzare solo la prima pagina del report. Questa opzione è utile per visualizzare in anteprima il formato del report prima di generarlo.

I dati sono già ordinati. Se il report include variabili di separazione, prima di generarlo è necessario ordinare il file dati in base ai valori delle variabili di separazione. Se il file di dati è già ordinato in base ai valori delle variabili di separazione, è possibile ridurre i tempi di elaborazione selezionando questa opzione. Questa opzione risulta particolarmente utile dopo aver visualizzato un'anteprima del report.

Per ottenere un riepilogo: Riepiloghi per colonne

- Dai menu, scegliere:
  Analizza
  Report
  Report: Riepiloghi per colonne...

- Selezionare una o più variabili per Colonne dati. Per ogni variabile selezionata verrà generata una colonna nel report.

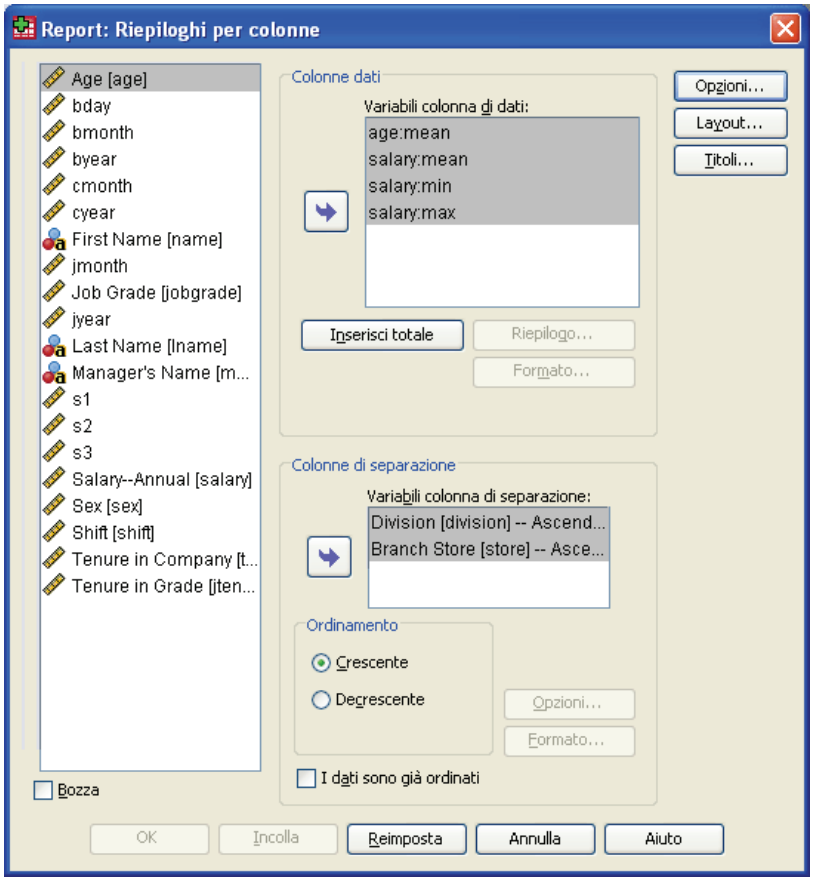
- Per modificare la misura riassuntiva relativa a una variabile, selezionare la variabile nell'elenco Variabili colonna di dati e fare clic su Riepilogo.

- Per ottenere più di una misura riassuntiva per una variabile, selezionare la variabile nell'elenco sorgente e spostarla nell'elenco Variabili colonna di dati più volte, una per ogni misura riassuntiva desiderata.

- Per visualizzare una colonna contenente la somma, la media, il rapporto o altre funzioni per le colonne esistenti, fare clic su Inserisci totale. In tal modo verrà inserita una variabile denominata *totale* nell'elenco Colonne dati.

- Per i report ordinati e visualizzati in base ai sottogruppi, selezionare una o più variabili per Colonne di separazione.

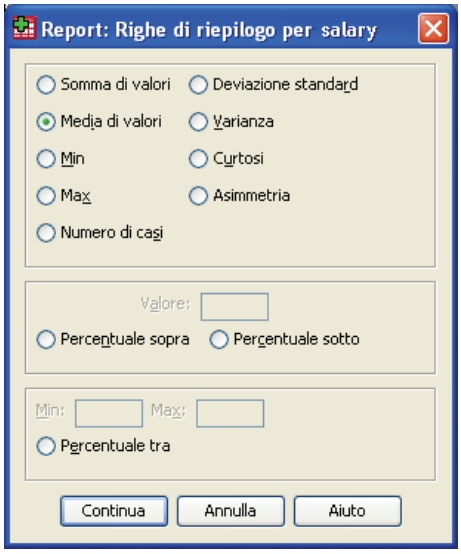
Figura 41-8
Finestra di dialogo Report: Riepiloghi per colonne



Funzione di rappresentazione delle colonne di dati

L'opzione Linee riassuntive consente di controllare le statistiche riassuntive visualizzate per la variabile della colonna dati selezionata.

Figura 41-9
Finestra di dialogo Report: Linee riassuntive



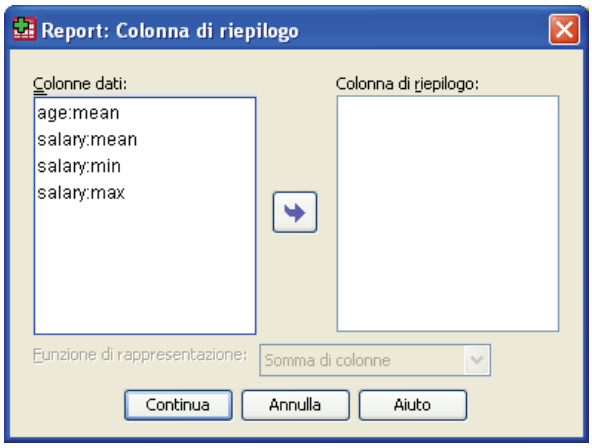
Le statistiche riassuntive disponibili sono: somma, media, minimo, massimo, numero di casi, percentuale di casi al di sopra o al di sotto di un valore specifico, percentuale di casi entro un intervallo specifico di valori, deviazione standard, varianza, curtosi e asimmetria.

Colonna di riepilogo del totale generale

L'opzione Colonna di riepilogo consente di gestire le statistiche riassuntive generali che riassumono due o più colonne dati.

Le statistiche riassuntive generali disponibili sono somma di colonne, media di colonne, minimo, massimo, differenza tra valori in due colonne, quoziente dei valori in una colonna divisi per i valori in un'altra colonna e prodotto di valori di colonne moltiplicati insieme.

Figura 41-10
Finestra di dialogo Report: Colonna di riepilogo



Somma di colonne. La colonna *totale* rappresenta la somma delle colonne nell'elenco Colonna di riepilogo.

Media di colonne. La colonna *totale* rappresenta la media delle colonne nell'elenco Colonna di riepilogo.

Minimo di colonne. La colonna *totale* rappresenta il minimo delle colonne nell'elenco Colonna di riepilogo.

Massimo di colonne. La colonna *totale* rappresenta il massimo delle colonne nell'elenco Colonna di riepilogo.

1a colonna – 2a colonna. La colonna *totale* rappresenta la differenza delle colonne nell'elenco Colonna di riepilogo. L'elenco Colonna di riepilogo deve contenere esattamente due colonne.

1a colonna / 2a colonna. La colonna *totale* rappresenta il quoziente delle colonne nell'elenco Colonna di riepilogo. L'elenco Colonna di riepilogo deve contenere esattamente due colonne.

% 1a colonna / 2a colonna. La colonna *totale* rappresenta la percentuale della prima colonna rispetto alla seconda colonna nell'elenco Colonna di riepilogo. L'elenco Colonna di riepilogo deve contenere esattamente due colonne.

Prodotto di colonne. La colonna *totale* rappresenta il prodotto delle colonne nell'elenco Colonna di riepilogo.

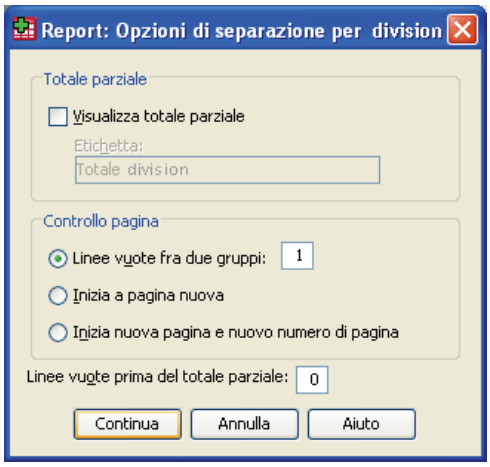
Formato delle colonne del report

Le opzioni di formattazione delle colonne dati e di separazione per i report di riepilogo per colonne sono le stesse descritte per i report di riepilogo per righe.

Report: Opzioni di separazione (Riepiloghi per colonne)

La funzione Opzioni di separazione consente di impostare la visualizzazione, la spaziatura e l'impaginazione per i gruppi.

Figura 41-11
Finestra di dialogo Report: Opzioni di separazione



Totale parziale. Consente di impostare la visualizzazione di totali parziali per i gruppi.

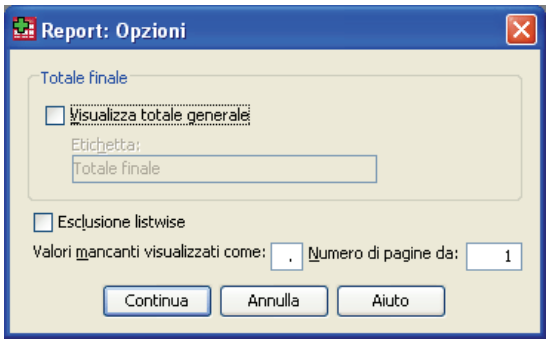
Controllo pagina. Consente di impostare la spaziatura e l'impaginazione per le categorie relative alla variabile di separazione selezionata. È possibile impostare il numero desiderato di righe vuote tra i gruppi o fare in modo che ciascun gruppo inizi in una nuova pagina.

Linee vuote prima del totale parziale. Consente di impostare il numero di righe vuote tra i dati dei gruppi e i totali parziali.

Report: Opzioni (Riepiloghi per colonne)

Le opzioni consentono di impostare la visualizzazione dei totali generali e dei valori mancanti e l'impaginazione dei report di riepilogo per colonne.

Figura 41-12
Finestra di dialogo Report: Opzioni



Totale finale. Consente di visualizzare ed etichettare un totale finale per ogni colonna, visualizzato alla fine della colonna.

Valori mancanti. È possibile escludere i valori mancanti dal report o selezionare un solo carattere per indicare i valori mancanti nel report.

Report: Layout per riepiloghi per colonne

Le opzioni di layout per i report di riepilogo per colonne sono le stesse descritte per i report di riepilogo per righe.

Funzioni aggiuntive del comando REPORT

Il linguaggio della sintassi dei comandi consente inoltre di:

- Visualizzare diverse funzioni di riepilogo nelle colonne di una linea riassuntiva.
- Inserire linee riassuntive nelle colonne dati per le variabili diverse dalle variabili della colonna, o per le varie combinazioni (funzioni composte) di funzioni di rappresentazione.
- Utilizzare Mediana, Moda, Frequenza e Percentuale come funzioni di rappresentazione.
- Controllare in modo più preciso il formato di visualizzazione delle statistiche riassuntive.

- Inserire linee vuote in corrispondenza di vari punti nei report.
- Inserire linee vuote dopo ogni *n* caso nei listati.

A causa della complessità della sintassi `REPORT`, può risultare utile, durante la costruzione di un nuovo report con la sintassi, approssimare il report generato dalle finestre di dialogo, copiare e incollare la sintassi corrispondente e ridefinire tale sintassi in modo da farla corrispondere al report specifico.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
42

Analisi di affidabilità

L'analisi di affidabilità consente di studiare le proprietà delle scale di misurazione e degli elementi che le compongono. La procedura Analisi di affidabilità calcola una serie di misure comunemente utilizzate in relazione all'affidabilità della scala e fornisce inoltre informazioni relative alle relazioni tra singoli elementi della scala. I coefficienti di correlazione tra classi possono essere utilizzati per calcolare le stime di affidabilità.

Esempio. Il questionario misura la soddisfazione del cliente in un modo utile? Utilizzando l'analisi di affidabilità, è possibile determinare il grado di correlazione tra gli elementi del questionario, ottenere un indice globale della ripetibilità oppure la concordanza interna della scala in modo globale. È quindi possibile identificare gli elementi del problema che devono essere esclusi dalla scala.

Statistiche. Descrittive per ogni variabile e per la scala, statistiche riassuntive degli elementi, correlazioni e covarianze tra elementi, stime di affidabilità, tabella ANOVA, coefficienti di correlazione tra classi, $T^{quadrato}$di Hotelling e test di additività di Tukey.

Modelli. Sono disponibili i seguenti modelli di affidabilità:

- **Alfa (Cronbach).** È un modello di concordanza interna, basato sulla media di correlazione fra elementi.
- **Divisione a metà.** Questo modello divide la scala in due parti ed esamina la correlazione tra le parti.
- **Guttman.** Questo modello calcola i limiti inferiori di Guttman per una reale affidabilità.
- **Parallelo.** Questo modello presume che tutti gli elementi abbiano varianze e varianze di errore uguali tra le replicazioni.
- **Parallelo esatto.** Questo modello afferma le ipotesi del modello parallelo e assume inoltre medie uguali degli elementi.

Dati. I dati possono essere dicotomici, ordinali oppure intervalli ma devono essere codificati numericamente.

Assunzioni. Le osservazioni devono essere indipendenti e gli errori non devono essere correlati agli elementi. Ogni coppia di elementi deve avere una distribuzione normale bivariata. Le scale devono essere additive, in modo che ogni elemento sia correlato in modo lineare al punteggio totale.

Procedure correlate. Se si desidera esplorare la dimensionalità degli elementi della scala (per verificare se è necessaria più di una costruzione per tenere conto del modello dei punteggi degli elementi), utilizzare la procedura Analisi fattoriale o Scaling multidimensionale. Per identificare i gruppi omogenei di variabili, usare la cluster gerarchica per raggruppare le variabili.

Per ottenere l'analisi di affidabilità

- Dai menu, scegliere:

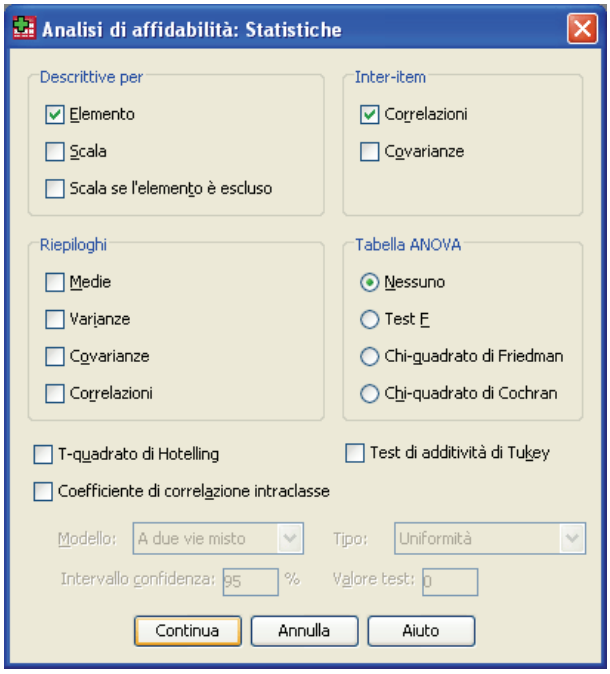
Analizza
  Scala
    Analisi di affidabilità...

Figura 42-1
Finestra di dialogo Analisi di affidabilità

- Selezionare due o più variabili come potenziali componenti di una scala additiva.
- Scegliere un modello dall'elenco a discesa Modello.

Analisi di affidabilità: Statistiche

Figura 42-2
Finestra di dialogo Analisi di affidabilità: Statistiche



È possibile selezionare varie statistiche per la descrizione della scala e degli elementi. Le statistiche riportate per impostazione predefinita comprendono il numero di casi, il numero di elementi e le stime di affidabilità riportati di seguito:

- **Modelli Alfa.** Per i dati dicotomici, è equivalente al coefficiente Kuder-Richardson 20 (KR20).
- **Modelli Divisione a metà.** Correlazione tra stime dei parametri, affidabilità di Divisione a metà di Guttman, affidabilità di Spearman-Brown (lunghezza uguale e diversa) e coefficiente alfa per ogni metà.
- **Modelli di Guttman.** Coefficienti di affidabilità da lambda 1 a lambda 6.
- **Modelli Parallelo e Parallelo esatto.** Test sulla bontà dell'adattamento del modello, stime della varianza di errore, matrice di correlazione e troncamento, media comune stimata, affidabilità stimata e dati non troncati.

Descrittive per. Fornisce statistiche descrittive per le scale o per gli elementi tra i casi.

- **Elemento.** Fornisce statistiche descrittive per gli elementi tra i casi.
- **Scala.** Fornisce statistiche descrittive per le scale.
- **Scala se l'item è escluso.** Consente di visualizzare statistiche riassuntive per il confronto di ogni elemento con la scala composta dagli altri elementi. Le statistiche includono la media della scala e la varianza risultante se l'elemento venisse eliminato dalla scala, la correlazione tra l'elemento e la scala composta dagli altri elementi e l'Alfa di Cronbach risultante se l'elemento venisse eliminato dalla scala.

Riepiloghi. Fornisce statistiche descrittive della distribuzione di elementi tra tutti gli elementi nella scala.

- **Medie.** Statistiche riassuntive sulle medie degli item. Vengono visualizzate la media più piccola, la più grande e la media centrale, nonché l'intervallo e la deviazione standard delle medie e il rapporto fra la media più grande e la più piccola.
- **Varianze.** Statistiche riassuntive per le varianze degli elementi. Vengono riprodotte la varianza minima, massima e media, nonché l'intervallo e il rapporto fra varianza massima e minima.
- **Covarianze.** Statistiche riassuntive basate sulle correlazioni tra item. Vengono visualizzate la covarianza più piccola, la più grande e la covarianza media, nonché l'intervallo e la deviazione standard delle covarianze e il rapporto fra la covarianza più grande e la più piccola.
- **Correlazioni.** Statistiche riassuntive basate sulle correlazioni tra item. Vengono visualizzate la correlazione più piccola, la più grande e la correlazione media, nonché l'intervallo e la deviazione standard delle correlazioni e il rapporto fra la correlazione più grande e la più piccola.

Inter-item. Fornisce matrici di correlazioni o covarianze tra elementi.

Tabella ANOVA. Fornisce test di medie uguali.

- **Test F.** Visualizza una tabella di analisi della varianza a misure ripetute.
- **Chi-quadrato di Friedman.** Visualizza il chi-quadrato di Friedman e il coefficiente di concordanza di Kendall. Questa opzione è appropriata per dati che rappresentano classifiche (ranghi). Il test chi-quadrato sostituisce il test F solitamente utilizzato nelle tabelle ANOVA.
- **Chi-quadrato di Cochran.** Visualizza la Q di Cochran. Questa opzione è adatta ai dati dicotomici. La Q di Cochran sostituisce il test F solitamente utilizzato nelle tabelle ANOVA.

T quadrato di Hotelling. Crea un test multivariato dell'ipotesi nulla in base alla quale tutti gli elementi sulla scala hanno la stessa media.

Test di additività di Tukey. Crea un test dell'ipotesi in base alla quale non esiste un'interazione moltiplicativa tra gli elementi.

Coefficiente di correlazione intraclasse. Crea misurazioni della consistenza o dell'accordo dei valori all'interno dei casi.

- **Modello.** Consente di selezionare il modello per calcolare i coefficienti di correlazione intraclasse. I modelli disponibili sono A due vie misto, A due vie casuale e A una via casuale. Selezionare A due vie misto quando gli effetti relativi alle persone sono casuali e quelli relativi all'elemento sono fissi, A due vie casuale quando sia gli effetti relativi alle persone che quelli relativi all'elemento sono casuali oppure A una via casuale quando gli effetti relativi alle persone sono casuali.
- **Tipo.** Consente di selezionare il tipo di indice. I tipi disponibili sono Uniformità e Concordanza assoluta.
- **Intervallo di confidenza.** Consente di specificare il livello relativo all'intervallo di confidenza. Il valore predefinito è il 95%.
- **Valore test.** Consente di specificare il valore ipotizzato del coefficiente relativo al test dell'ipotesi. Si tratta del valore rispetto al quale viene confrontato il valore osservato. Il valore predefinito è 0.

Opzioni aggiuntive del comando RELIABILITY

Il linguaggio della sintassi dei comandi consente inoltre di:

- Leggere ed analizzare una matrice di correlazione.
- Scrivere una matrice di correlazione per analizzarla in seguito.
- Specificare le divisioni diverse dalle metà uguali per il metodo di divisione a metà.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
43

Scaling multidimensionale

La procedura Scaling multidimensionale consente di effettuare un tentativo per trovare la struttura in un insieme di misure di distanza tra oggetti o casi. Questa operazione viene compiuta assegnando le osservazioni a posizioni specifiche in uno spazio concettuale (in genere bi o tridimensionale) in modo che le distanze tra i punti nello spazio corrispondano il più possibile alle dissimilarità specificate. In molti casi, le dimensioni di questo spazio concettuale possono essere interpretate ed utilizzate allo scopo di comprendere meglio i dati.

Se si dispone di variabili misurate oggettivamente, è possibile utilizzare lo scaling multidimensionale come una tecnica di riduzione dei dati (la procedura calcolerà le distanze dai dati multivariati per l'utente, se necessario). È inoltre possibile applicare lo scaling multidimensionale alle stime soggettive di dissimilarità tra oggetti o concetti. In aggiunta, Scaling multidimensionale può gestire i dati di dissimilarità da più origini, come nel caso di più stimatori o di rispondenti ai questionari.

Esempio. In che modo i consumatori percepiscono le similitudini tra automobili diverse? Se si dispone di dati che rilevano similarità tra diverse forme e modelli di automobili, lo scaling multidimensionale consentirà di identificare le dimensioni in grado di descrivere le percezioni dei consumatori. È possibile verificare, ad esempio, che il prezzo e le dimensioni di un veicolo definiscono uno spazio bidimensionale, secondo le spiegazioni fornite dai rispondenti.

Statistiche. Per ciascun modello: matrice dei dati, matrice di dati scalati ottimale, s-stress (di Young), stress (di Kruskal), RSQ, coordinate degli stimoli, media dello stress e RSQ per ogni stimolo (modelli RMDS). Per modelli di singole differenze (INDSCAL): pesi di soggetto e valore non ponderato per ogni soggetto. Per ogni matrice in modelli di scaling multidimensionale replicati: stress e RSQ per ogni stimolo. Grafici: coordinate degli stimoli (bi o tridimensionali), grafico a dispersione delle disparità rispetto alle distanze.

Dati. Se si dispone di dati di dissimilarità, tutte le dissimilarità dovrebbero essere quantitative e dovrebbero essere misurate in base alla stessa metrica. Se i dati sono multivariati, le variabili possono essere quantitative, binarie o dati di conteggio. Lo scaling delle variabili è una questione importante poiché le differenze possono influenzare la soluzione. Se le variabili hanno differenze significative (ad esempio, una variabile è misurata in dollari e l'altra è misurata in anni), è consigliabile standardizzarle. Questa operazione può essere eseguita automaticamente dalla procedura Scaling multidimensionale.

Assunzioni. La procedura Scaling multidimensionale è relativamente libera da ipotesi di distribuzione. Assicurarsi di selezionare il livello di misurazione appropriato (ordinale, intervallo o rapporto) nella finestra di dialogo Scaling multidimensionale per essere sicuri che i risultati vengano calcolati correttamente.

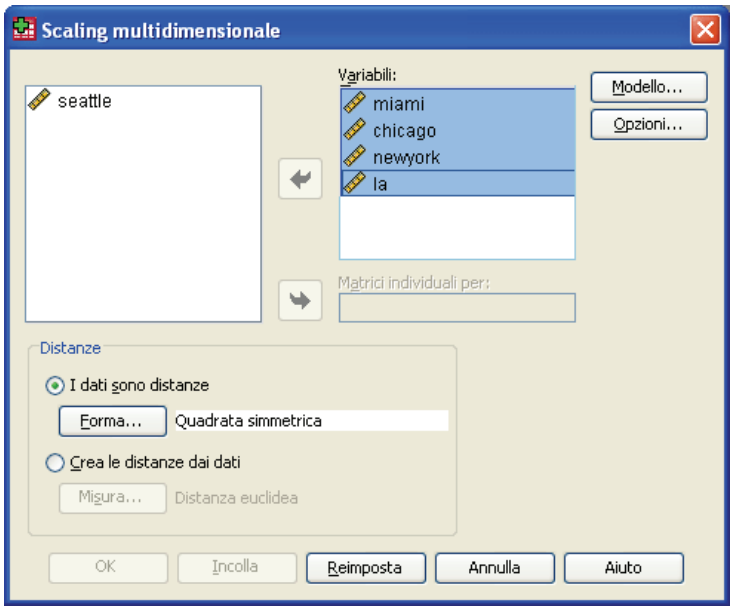
Procedure correlate. Se l'obiettivo è la riduzione dei dati, si può considerare un metodo alternativo quale l'analisi fattoriale, in particolare se le variabili sono di tipo quantitativo. Se si desidera identificare gruppi di casi simili, considerare la possibilità di integrare l'analisi di scaling multidimensionale con un'analisi gerarchica o cluster *k*-medie.

Per ottenere un'analisi Scaling multidimensionale

- Dai menu, scegliere:

Analizza
  Scala
    Scaling multidimensionale...

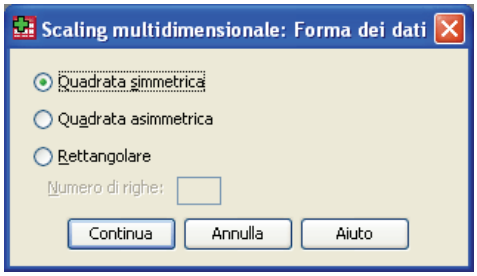
Figura 43-1
Finestra di dialogo Scaling multidimensionale



- Nel gruppo Distanze selezionare I dati sono distanze oppureCrea le distanze dai dati.
- Se i dati sono rappresentati da distanze, selezionare almeno quattro variabili numeriche per l'analisi. È possibile anche fare clic su Forma per specificare la forma della matrice delle distanze.
- Se si desidera creare le distanze prima di analizzarle, selezionare almeno una variabile numerica. È possibile anche fare clic su Misura per specificare il tipo di misura di distanza desiderato. Ciascuna variabile di raggruppamento può essere un numero o una stringa ed è possibile creare matrici diverse per ciascuna categoria di variabili spostando la variabile in questione nell'elenco Matrici individuali per.

Scaling multidimensionale: Forma dei dati

Figura 43-2
Finestra di dialogo Forma scaling multidimensionale

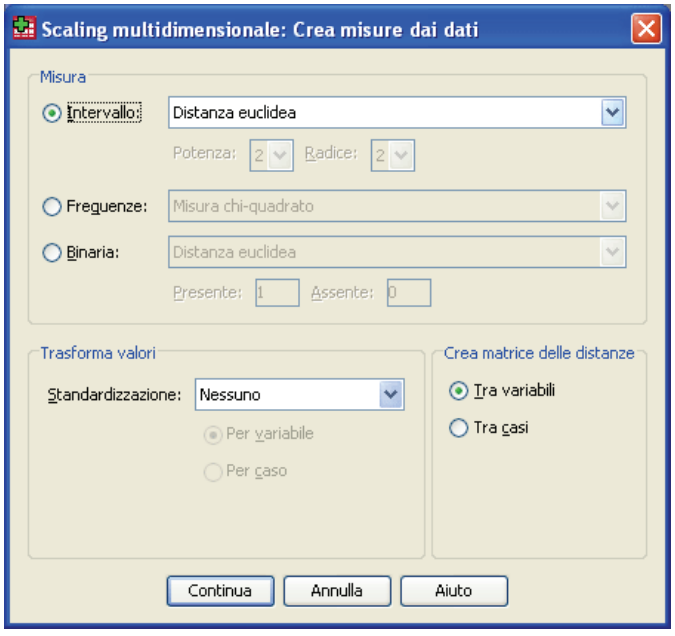


Se il file dati attivo rappresenta le distanze tra uno o due insiemi di oggetti, è necessario specificare la forma della matrice dei dati in modo da ottenere i risultati corretti.

Nota: Non è possibile selezionare Quadrata simmetrica se la finestra di dialogo Modello specifica la riga in modo condizionale.

Scaling multidimensionale: Crea misure dai dati

Figura 43-3
Finestra di dialogo Scaling multidimensionale: Crea misure dai dati



La procedura Scaling multidimensionale utilizza dati di dissimilarità per creare una soluzione di scaling. Se i dati disponibili sono dati multivariati (valori di variabili misurate), è necessario creare dati di dissimilarità in modo da calcolare una soluzione di scaling multidimensionale. È possibile specificare i dettagli della creazione delle misure di dissimilarità a partire dai dati disponibili.

Misura. Consente di specificare la misura di dissimilarità per l'analisi. Selezionare un'alternativa dal gruppo Misura corrispondente al tipo di dati desiderato e quindi selezionare una delle misure dall'elenco a discesa corrispondente a tale tipo di misura. Le alternative disponibili sono:

- **Intervallo.** Distanza euclidea, Distanza euclidea quadratica, Chebychev, City-Block, Minkowski o Personalizzato.
- **Conteggi.** Misura chi-quadrato e Misura phi-quadrato.
- **Binaria.** Distanza euclidea, Distanza euclidea quadratica, Differenza di dimensione, Differenza di modello, Varianza o Lance e Williams.

Crea matrice delle distanze. Consente di scegliere l'unità di analisi. Le alternative sono Fra variabili o Fra casi.

Trasforma valori. In alcuni casi, ad esempio quando le variabili sono misurate su scale molto diverse, è possibile standardizzarne i valori prima di calcolare le dissimilarità (non applicabile ai dati binari). Selezionare un metodo di standardizzazione dall'elenco a discesa Standardizza. Se non è richiesta alcuna standardizzazione, selezionare Nessuna.

Scaling multidimensionale: Modello

Figura 43-4
Finestra di dialogo Scaling multidimensionale: Modello

La stima corretta di un modello di scaling multidimensionale dipende dagli aspetti dei dati e del modello stesso.

Livello di misurazione. Consente di specificare il livello dei dati. Le alternative sono Ordinale, Intervallo o Rapporto. Se le variabili sono ordinali, la selezione dell'opzione Distingui le osservazioni pari merito richiede che vengano trattate come variabili continue, in modo che i pari merito (valori uguali per casi differenti) siano risolti in modo ottimale.

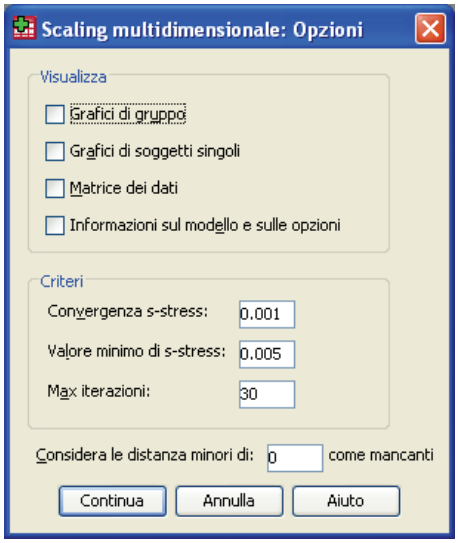
Condizionalità. Consente di specificare quali confronti sono significativi. Le alternative sono Matrice, Riga o Non condizionale.

Dimensioni. Consente di specificare la dimensione della soluzione di scaling. Per ciascun numero nell'intervallo viene calcolata una soluzione. Specificare interi da 1 a 6. È consentito un minimo di 1 solo se si è selezionato Distanza euclidea come modello di scaling. Per una soluzione singola, specificare lo stesso valore minimo e massimo.

Modello di scaling. Consente di specificare le ipotesi in base alle quali viene eseguito lo scaling. Le alternative disponibili sono Distanza euclidea o Distanza euclidea per differenze singole (nota anche come INDSCAL). Nel modello Distanza euclidea per differenze singole, è possibile selezionare l'opzione Accetta pesi di soggetto negativi, se appropriata per i dati disponibili.

Scaling multidimensionale: Opzioni

Figura 43-5
Finestra di dialogo Scaling multidimensionale: Opzioni



È possibile impostare le opzioni per l'analisi scaling multidimensionale:

Visualizzazione. Consente di selezionare vari tipi di output. Le opzioni disponibili sono Grafici di gruppo, Grafici di soggetti singoli, Matrice dei dati e Informazioni sul modello e sulle opzioni.

Criteri. Consente di determinare il momento in cui interrompere l'iterazione. Per modificare i valori predefiniti, inserire i valori per Convergenza s-stress, Valore minimo di s-stress e Massimo numero di iterazioni.

Considera le distanze minori di N come mancanti. Le distanze minori di tale valore sono escluse dall'analisi.

Opzioni aggiuntive del comando ALSCAL

Il linguaggio della sintassi dei comandi consente inoltre di:

- Utilizzare tre tipi di modelli aggiuntivi, noti come ASCAL, AINDS e GEMSCAL, per lo scaling multidimensionale.
- Eseguire trasformazioni polinomiali su dati di intervallo e di rapporto.

- Analizzare le similarità (piuttosto che le distanze) con dati ordinali.
- Analizzare i dati nominali.
- Salvare varie matrici di coordinate e pesi nei file e rileggerli per l'analisi.
- Vincolare l'unfolding multidimensionale.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Capitolo
44

Statistiche di rapporto

La procedura Statistiche di rapporto offre un elenco completo di statistiche riassuntive per la descrizione del rapporto tra due variabili di scala.

È possibile ordinare l'output in base ai valori di una variabile di raggruppamento in ordine crescente o decrescente. È possibile eliminare il report sulle statistiche di rapporto dall'output e salvare i risultati in un file esterno.

Esempio. Esiste un buon grado di uniformità nel rapporto tra prezzo di stima e prezzo di vendita delle case in ognuno dei cinque paesi? Dall'output, si apprende che la distribuzione dei rapporti varia notevolmente da paese a paese.

Statistiche. Mediana, media, media pesata, intervalli di confidenza, coefficiente di dispersione, coefficiente di variazione centrato sulla mediana, coefficiente di variazione centrato sulla media, differenziale di prezzo, deviazione standard, deviazione assoluta media, intervallo, valori minimo e massimo e indice di concentrazione calcolato per una percentuale o un intervallo specifico per l'utente compresi nel rapporto della mediana.

Dati. Utilizzare codici numerici o stringhe per codificare le variabili di raggruppamento (misure di livello nominale o ordinale).

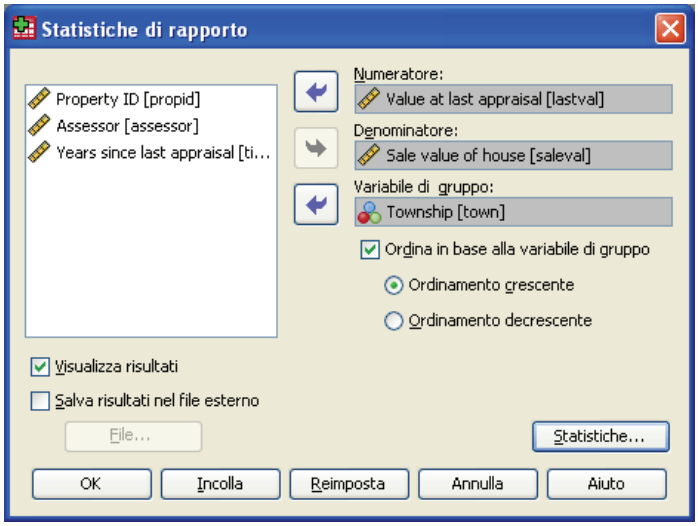
Assunzioni. Le variabili che definiscono il numeratore e il denominatore del rapporto dovrebbero essere variabili di scala con valori positivi.

Per ottenere statistiche di rapporto

- Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Rapporto...

Figura 44-1
Finestra di dialogo Statistiche di rapporto



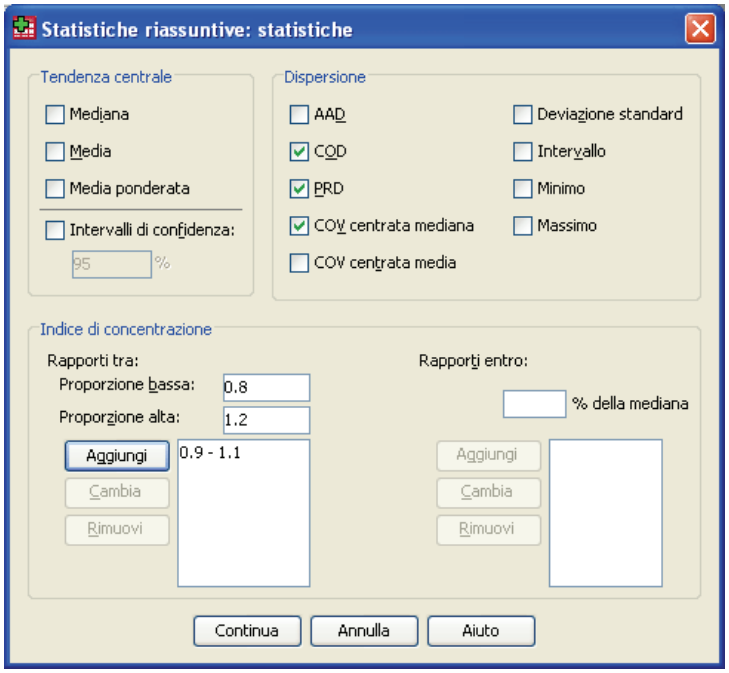
- Selezionare una variabile numeratore.
- Selezionare una variabile denominatore.

Oppure:

- Selezionare una variabile di raggruppamento e specificare l'ordinamento dei gruppi nei risultati.
- Scegliere se visualizzare o meno i risultati nel Viewer.
- Decidere se salvare o meno i risultati in un file esterno per futuri utilizzi. In caso affermativo, specificare il nome del file in cui verranno salvati.

Statistiche di rapporto

Figura 44-2
Finestra di dialogo Statistiche di rapporto



Tendenza centrale. Le misure di tendenza centrale sono statistiche che descrivono la distribuzione dei rapporti.

- **Mediana.** Il valore tale che il numero di rapporti inferiore a questo valore e il numero di rapporti maggiore di questo valore siano uguali.
- **Media.** Il risultato della somma dei rapporti e la divisione del risultato per il numero totale di rapporti.
- **Media ponderata.** Il risultato della divisione della media del numeratore per la media del denominatore. La media pesata è anche la media dei rapporti ponderati dal denominatore.
- **Intervalli di confidenza.** Vengono visualizzati gli intervalli di confidenza per la media, la mediana e la media ponderata (se necessario). Specificare un valore maggiore o uguale a 0 e minore di 100 come intervallo di confidenza.

Dispersione. Queste statistiche misurano il grado di variazione o variabilità nei valori osservati.

- **AAD.** (Deviazione assoluta media) È ottenuta sommando le deviazioni assolute dei rapporti dalla mediana e dividendo il risultato per il numero totale di rapporti.
- **COD.** (Coefficiente di dispersione) È il risultato dell'espressione della deviazione assoluta media come una percentuale della mediana.
- **PRD.** (Differenziale di prezzo) Anche noto come indice di regressività, è il risultato della divisione della media per la media pesata.
- **COV centrata mediana.** (Coefficiente di variazione) È il risultato dell'espressione delle radici quadrate medie della deviazione dalla mediana come una percentuale della mediana.

- **COV centrata media.** (Coefficiente di variazione) È il risultato dell'espressione della deviazione standard come una percentuale della media.
- **Deviazione standard.** La deviazione standard viene ottenuta sommando le deviazioni quadrate dei rapporti dalla media, dividendo il risultato per il numero totale di rapporti meno uno ed estraendo la radice quadrata positiva.
- **Intervallo.** L'intervallo è il risultato della differenza tra rapporto massimo e rapporto minimo.
- **Minimo.** Il minimo è il rapporto più piccolo.
- **Massimo.** Il massimo è il rapporto più grande.

Indice di concentrazione. Consente di misurare la percentuale di rapporti che cade in un intervallo. È possibile calcolarlo in due modi diversi:

- **Rapporti tra.** In questo caso l'intervallo viene definito in modo esplicito specificando i valori alti e bassi dell'intervallo. Immettere i valori per le proporzioni alte e basse, quindi fare clic su Aggiungi per ottenere un intervallo.
- **Rapporti entro.** In questo caso l'intervallo viene definito in modo implicito specificando la percentuale della mediana. Digitare un valore compreso tra 0 e 100 e fare clic su Aggiungi. Il valore minimo dell'intervallo è uguale a (1 – 0,01 × valore) × mediana, e il massimo è uguale a (1 + 0,01 × valore) × mediana.

Capitolo
45

Curva ROC

Questa procedura è utile per valutare le prestazioni degli schemi di classificazione in cui i soggetti sono classificati in base a una variabile con due categorie.

Esempio. Poiché è interesse della banca classificare in modo corretto i clienti adempienti o meno, è necessario elaborare metodi specifici per decidere a chi concedere i prestiti. Le curve ROC possono essere utilizzate per valutare il livello di attendibilità di questi metodi.

Statistiche. Area sotto alla curva ROC con intervallo di confidenza e coordinate della curva ROC. Grafici: Curva ROC.

Metodi. È possibile calcolare una stima dell'area sotto alla curva ROC in modo non parametrico o parametrico mediante un modello esponenziale binegativo.

Dati. Le variabili del test sono quantitative. Le variabili del test sono spesso composte dalle probabilità dell'analisi discriminante o della regressione logistica oppure da punteggi su scala arbitraria, i quali rappresentano la "forza della convinzione" di chi classifica i soggetti in una categoria piuttosto che in un'altra. La variabile di stato può essere di qualsiasi tipo e indica la vera categoria a cui appartiene un soggetto. Il valore della variabile di stato indica quale categoria debba essere considerata *positiva*.

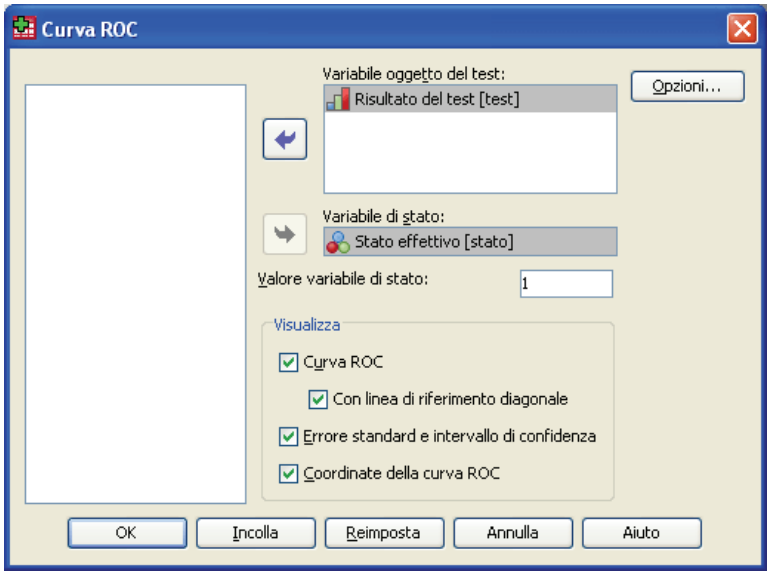
Assunzioni. Si suppone che i numeri crescenti sulla scala della classifica rappresentino la convinzione crescente che il soggetto appartenga a una categoria, mentre i numeri decrescenti rappresentino la convinzione crescente che il soggetto appartenga all'altra categoria. L'utente deve scegliere qual è la direzione *positiva*. Si suppone inoltre che la categoria *vero* alla quale appartiene ciascun soggetto sia conosciuta.

Per creare una Curva ROC

- Dai menu, scegliere:

  Analizza
  Curva ROC...

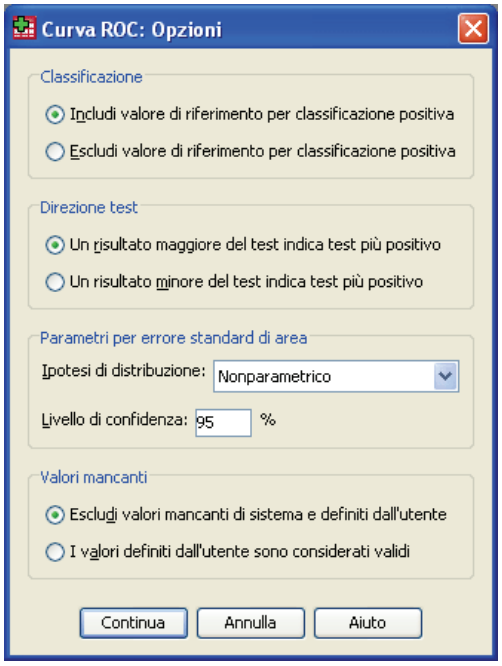
Figura 45-1
Finestra di dialogo Curva ROC



- Selezionare una o più variabili di probabilità per il test.
- Selezionare una variabile di stato.
- Individuare il valore *positivo* per la variabile di stato.

Curva ROC: Opzioni

Figura 45-2
Finestra di dialogo Curva ROC: Opzioni



È possibile selezionare le seguenti opzioni per l'analisi ROC:

Classificazione. Consente di specificare se il valore di riferimento debba essere incluso o escluso dalla classificazione *positiva*. Questa impostazione non ha attualmente alcun effetto sull'output.

Direzione test. Consente di specificare la direzione della scala in rapporto alla categoria *positiva*.

Parametri per errore standard di area. Consente di specificare il metodo di valutazione dell'errore standard dell'area sotto alla curva. I metodi disponibili sono non parametrico ed esponenziale binegativo. Consente inoltre di impostare il livello dell'intervallo di confidenza. L'intervallo disponibile è da 50,1% a 99,9%.

Valori mancanti. Consente di specificare le modalità di gestione dei valori mancanti.

Capitolo
46

Panoramica dello strumento Grafico

Per creare grafici ad alta risoluzione è possibile utilizzare le procedure disponibili nel menu Grafici e numerose procedure del menu Analizza. In questo capitolo viene fornita una panoramica dello strumento Grafico.

Creazione e modifica di un grafico

Prima di poter creare un grafico è necessario che i dati siano presenti nell'Editor dei dati. È possibile immettere i dati direttamente nell'Editor dei dati, aprire un file di dati salvato in precedenza, oppure leggere un foglio elettronico, un file di dati delimitato da tabulazioni o un file di database. Tramite il comando Esercitazione del menu ? è possibile accedere a esempi in linea relativi alla creazione e alla modifica di un grafico e nell'Aiuto sono disponibili informazioni sulla creazione e la modifica di tutti i tipi di grafico.

Creazione di grafici

Il Generatore di grafici permette di creare grafici da modelli di grafici predefiniti o da singole parti (ad esempio assi e barre) Per creare un grafico, è sufficiente trascinare i modelli di grafici o le singole parti nel **disegno** rappresentato dall'area di grandi dimensioni a destra dell'elenco Variabili nella finestra di dialogo Generatore di grafici.

L'area di disegno visualizza un'anteprima del grafico mentre è in corso la creazione. Sebbene l'anteprima utilizzi le etichette di variabile definite e i livelli di misurazione, non visualizza i dati effettivi. Utilizza invece dati generati casualmente per fornire una bozza approssimativa dell'aspetto del grafico.

L'uso dei modelli è il metodo di creazione più semplice da usare per i nuovi utenti. Per informazioni sull'uso dei modelli, vedere Creazione di un grafico dai modelli a pag. 506.

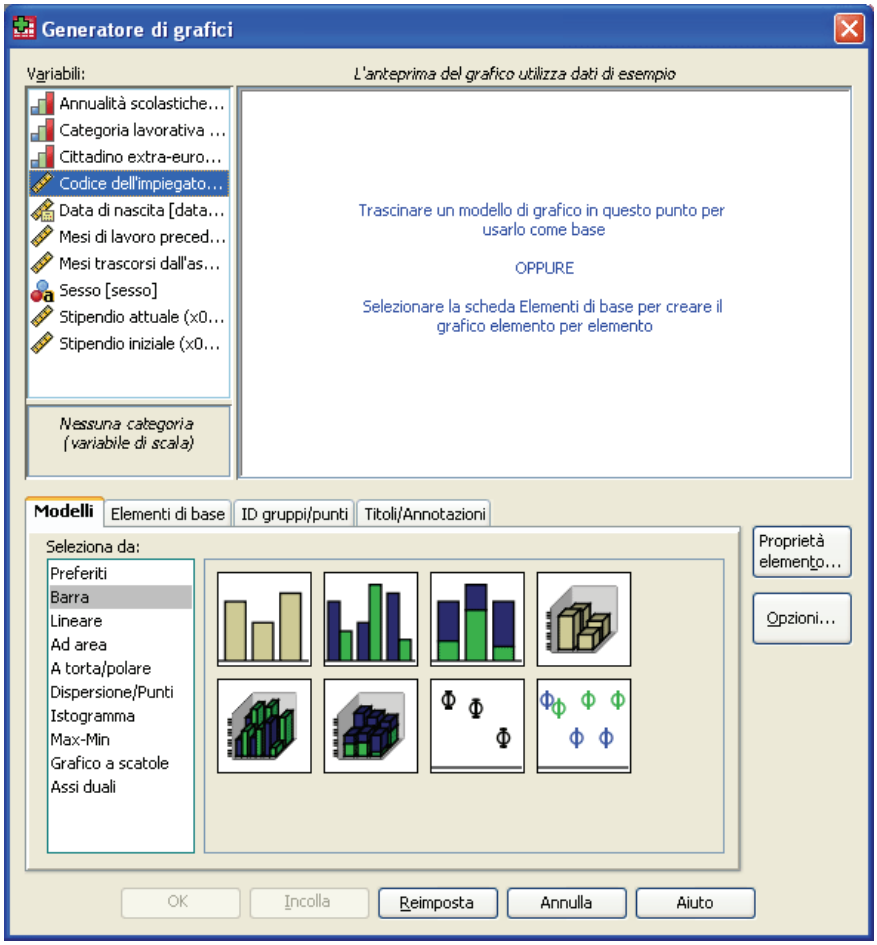
Per avviare il Generatore di grafici

▶ Dai menu, scegliere:

Grafici
  Generatore di grafici

Verrà visualizzata la finestra di dialogo corrispondente.

Figura 46-1
Finestra di dialogo Generatore di grafici



Creazione di un grafico dai modelli

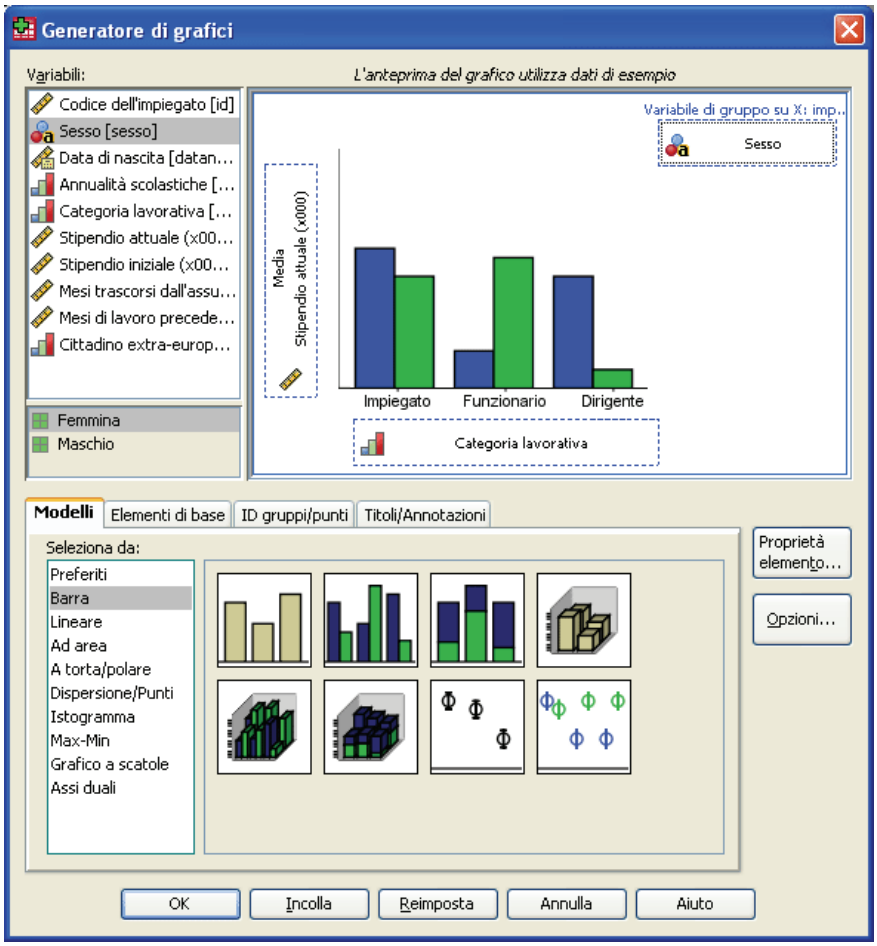
Il metodo più semplice per la creazione di grafici consiste nell'usare i modelli. Di seguito viene descritta la procedura generale per la creazione di un grafico dai modelli.

- Fare clic sulla scheda Modelli se non è già visualizzata.
- Nell'elenco Seleziona da, scegliere una categoria di grafici. Ciascuna categoria contiene vari tipi di grafici.
- Trascinare l'immagine del grafico desiderato nell'area del disegno. È possibile anche fare doppio clic sull'immagine. Se il disegno mostra già un grafico, il grafico modello sostituisce l'insieme di assi e gli elementi grafici del grafico.
- Trascinare le variabili desiderate dall'elenco Variabili nelle aree di rilascio dell'asse e se disponibile nell'area di rilascio del raggruppamento. Se l'area di rilascio degli assi visualizza già una statistica che si desidera utilizzare, non è necessario trascinare la variabile nell'area di

rilascio. La variabile deve essere inserita in un'area solo il testo dell'area è blu. Se il testo è nero, l'area contiene già una variabile o statistica.

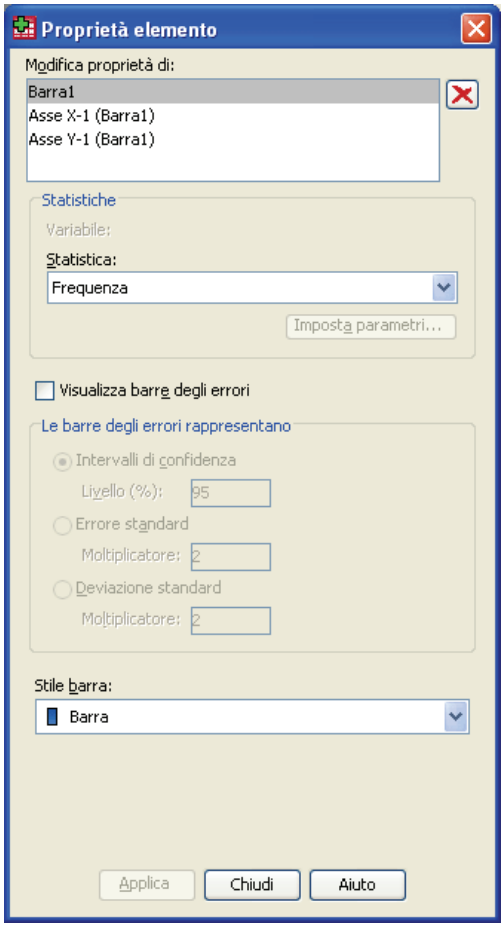
Nota: il livello di misurazione delle variabili è importante. Il Generatore di grafici imposta i valori predefiniti in base al livello di misurazione del grafico. Inoltre, il grafico risultante può differire in base ai livelli di misurazione usati. È possibile modificare temporaneamente il livello di misurazione di una variabile facendo clic con il pulsante destro del mouse su una variabile e selezionando un'opzione.

Figura 46-2
Finestra di dialogo Generatore di grafici con aree di rilascio compilate



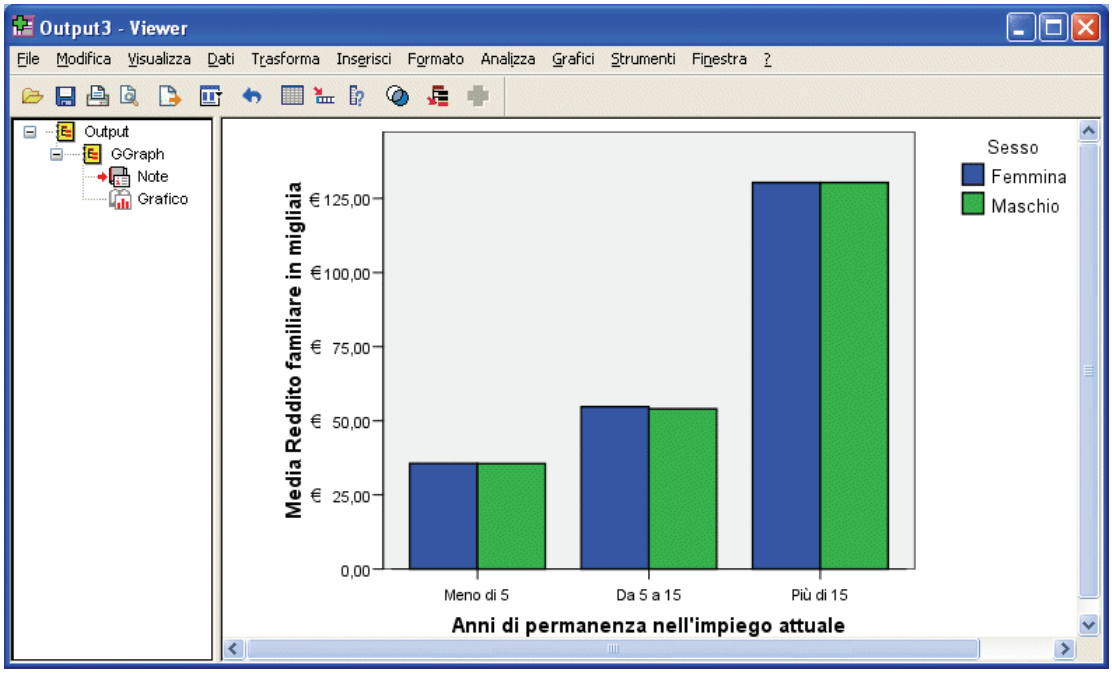
- Per modificare le statistiche o gli attributi degli assi o delle didascalie (ad esempio l'intervallo dei valori), fare clic su Proprietà elemento.

Figura 46-3
Finestra di dialogo Proprietà elemento



- Nell'elenco Modifica proprietà di, selezionare l'elemento da modificare. Per informazioni sulle proprietà specifiche, fare clic su Aiuto.
- Al termine delle modifiche, fare clic su Applica.
- Se è necessario aggiungere altre variabili al grafico (ad esempio per il raggruppamento o la definizione dei riquadri), fare clic sulla scheda ID gruppi/punti nella finestra di dialogo Generatore di grafici e selezionare una o più opzioni. Quindi trascinare le variabili categoriali nelle nuove aree di rilascio visualizzate nell'area di disegno.
- Per trasporre il grafico (ad esempio per convertire le barre in orizzontale), fare clic sulla scheda Elementi di base, quindi selezionare Trasponi.
- E fare clic su OK per creare il grafico. Il grafico viene visualizzato nel Viewer.

Figura 46-4
Grafico a barre visualizzato nella finestra del Viewer



Modifica dei grafici

L'Editor dei grafici offre un ambiente efficace e semplice da utilizzare per personalizzare i grafici e visualizzare i dati. L'Editor dei grafici offre le funzionalità seguenti:

- **Interfaccia utente semplice e intuitiva.** È possibile selezionare e modificare velocemente parti di grafici tramite menu, menu di scelta rapida e barre degli strumenti. È anche possibile immettere il testo direttamente nel grafico.
- **Ampia gamma di opzioni di formattazione e statistiche.** È possibile scegliere tra un'ampia gamma di stili e di opzioni statistiche.
- **Strumenti di esplorazione efficaci.** È possibile esplorare i dati in diversi modi, ad esempio tramite l'utilizzo delle etichette, il riordinamento o la rotazione. È possibile modificare i tipi di grafico e i ruoli delle variabili nel grafico. È inoltre possibile aggiungere curve di distribuzione nonché curve di adattamento, interpolazione o riferimento.
- **Modelli flessibili che consentono di ottenere un aspetto e un comportamento coerenti.** È possibile creare modelli personalizzati e utilizzarli per creare in modo semplice grafici con l'aspetto e le opzioni desiderate. Se ad esempio si desidera sempre un orientamento specifico per le etichette degli assi, è possibile definire tale orientamento in un modello e applicare il modello agli altri grafici.

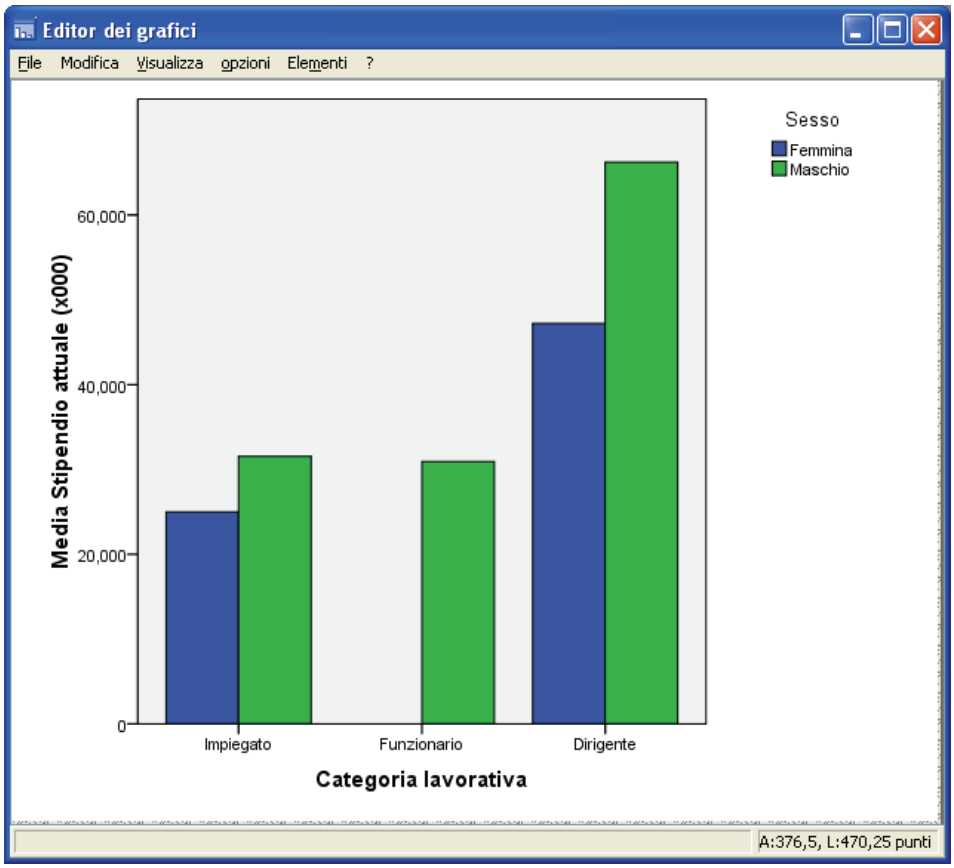
Per visualizzare l'Editor dei grafici

▶ Creare un grafico in SPSS Statistics oppure aprire un file Viewer contenente grafici.

▶ Fare doppio clic su un grafico nel Viewer.

Il grafico verrà visualizzato nell'Editor dei grafici.

Figura 46-5
Grafico visualizzato nell'Editor dei grafici



Nozioni fondamentali sull'Editor dei grafici

L'Editor dei grafici offre numerose modalità di gestione dei grafici.

Menu

Nell'Editor dei grafici molte operazioni vengono eseguite tramite i menu, in particolar modo quando si aggiunge un elemento a un grafico. Tramite i menu è ad esempio possibile aggiungere una curva di adattamento a un grafico a dispersione. Dopo avere aggiunto un elemento a un grafico è possibile specificare le relative opzioni tramite la finestra di dialogo Proprietà.

Finestra di dialogo Proprietà

Nella finestra di dialogo Proprietà sono disponibili le opzioni relative al grafico e agli elementi che lo compongono.

Per visualizzare la finestra di dialogo Proprietà è possibile:

- Fare doppio clic su un elemento del grafico.

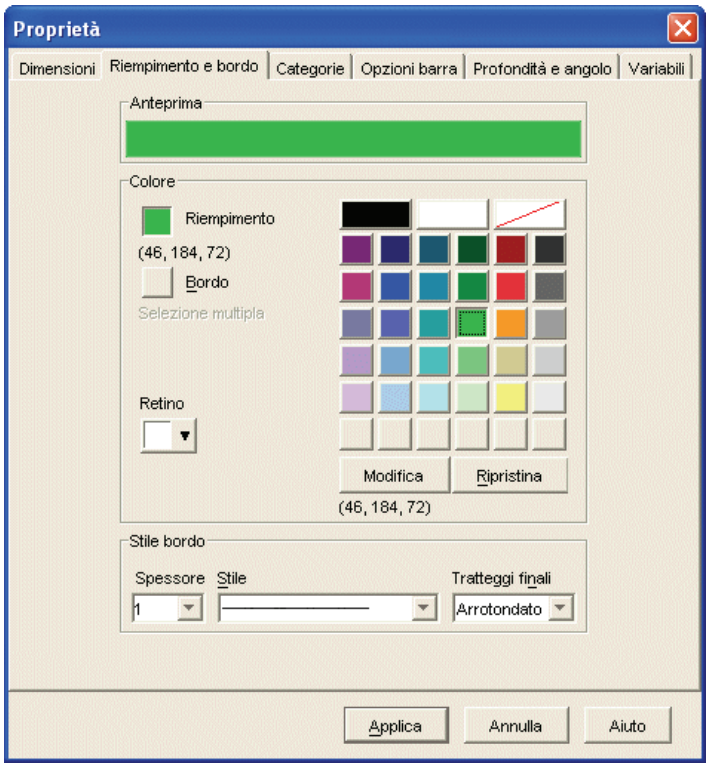
o un

- Selezionare un elemento del grafico e quindi dai menu, scegliere:

Modifica
Proprietà

In alternativa, la finestra di dialogo Proprietà viene visualizzata automaticamente quando si aggiunge un elemento al grafico.

Figura 46-6
Finestra di dialogo Proprietà, scheda Riempimento e bordo



La finestra di dialogo Proprietà è formata da schede che consentono di impostare le opzioni ed effettuare altre modifiche nei grafici. Le schede visualizzate nella finestra di dialogo Proprietà dipendono dall'elemento selezionato.

In alcune schede è disponibile un'anteprima che indica il risultato delle modifiche qualora vengano applicate all'elemento selezionato. Le modifiche verranno apportate al grafico solo dopo avere scelto Applica. Prima di scegliere Applica è possibile apportare modifiche in più di una scheda. Per cambiare la selezione e modificare un elemento diverso, fare clic su Applica prima di modificare la selezione. Se non si seleziona Applica prima di modificare la selezione e si seleziona Applica in seguito, le modifiche verranno applicate solo all'elemento o agli elementi selezionati.

In base all'elemento selezionato, solo alcune impostazioni potrebbero essere disponibili. La guida delle singole schede specifica gli elementi da selezionare per visualizzare le schede. Se vengono selezionati più elementi, sarà possibile modificare solo le impostazioni comuni a tutti gli elementi.

Barre degli strumenti

Le barre degli strumenti offrono un collegamento ad alcune delle funzionalità della finestra di dialogoProprietà. Invece di utilizzare la scheda Testo della finestra di dialogo Proprietà, è ad esempio possibile modificare il tipo di carattere e lo stile del testo tramite la barra degli strumenti Modifica.

Salvataggio delle modifiche

Le modifiche apportate al grafico vengono salvate quando si chiude l'Editor dei grafici. Il grafico modificato viene successivamente visualizzato nel Viewer.

Opzioni definizione grafico

Quando si definisce un grafico nel Generatore di grafici, è possibile aggiungere titoli e modificare le opzioni per la creazione dei grafici.

Aggiunta e modifica di titoli e annotazioni

È possibile aggiungere al grafico titoli e annotazioni per consentire a uno spettatore di interpretarlo più facilmente. Il Generatore di grafici mostra automaticamente nelle annotazioni le informazioni relative alla barra di errore.

Come aggiungere titoli e annotazioni

- Fare clic sulla scheda Titoli/Annotazioni.
- Selezionare uno o più titoli e annotazioni. Il disegno mostra del testo per indicare che i titoli e le annotazioni sono stati aggiunti al grafico.
- Utilizzare la finestra di dialogo Proprietà elemento per modificare il testo dei titoli e delle annotazioni.

Come rimuovere un titolo o un'annotazione

- Fare clic sulla scheda Titoli/Annotazioni.
- Deselezionare il titolo o l'annotazione che si desidera rimuovere.

Come modificare il testo del titolo o dell'annotazione

Quando si aggiungono titoli e annotazioni, non è possibile modificare il testo loro associato direttamente nel grafico. Come per altre voci del Generatore di grafici, per modificarli è necessario usare la finestra Proprietà elemento.

- Fare clic su Proprietà elementi se la finestra di dialogo corrispondente non è visualizzata.
- Nell'elenco Modifica proprietà di, selezionare un titolo, un sottotitolo o un'annotazione (ad esempio Titolo 1).
- Nella casella del contenuto, immettere il testo desiderato per il titolo, il sottotitolo o l'annotazione.
- Fare clic su Applica.

Impostazione delle opzioni generali

Il Generatore di grafici visualizza le opzioni generali disponibili per il grafico. Queste opzioni si applicano a tutto il grafico anziché a un elemento specifico. Le opzioni generali comprendono le opzioni per la gestione dei valori mancanti, i modelli, le dimensioni dei grafici e il ritorno a capo automatico nei riquadri.

- Fare clic su Opzioni.
- Modificare le opzioni generali. Vengono visualizzati i dettagli.
- Fare clic su Applica.

Valori mancanti definiti dall'utente

Variabili di separazione. Se ci sono valori mancanti nelle variabili usate per definire le categorie o i sottogruppi, selezionare Includi in modo che la categoria o le categorie dei valori mancanti definiti dall'utente (valori che sono stati identificati come mancanti dall'utente) vengano inclusi nel grafico. Queste categorie "mancanti" possono essere usate anche come variabili di separazione durante il calcolo di statistiche. Le categorie "mancanti" possono essere visualizzate sull'asse delle categorie o nelle didascalie, ad esempio aggiungendo una barra aggiunta o una fetta a un grafico a torta. Se non ci sono valori mancanti, le categorie "mancanti" non vengono visualizzate.

Se si seleziona questa opzione e si desidera disattivare la visualizzazione dopo che il grafico è stato creato, selezionare il grafico nell'Editor dei grafici, quindi scegliere Proprietà dal menu Modifica. Nella scheda Categorie, spostare le categorie che si desidera eliminare nell'elenco Esclusi. Le statistiche non vengono ricalcolate se le categorie "mancanti" vengono nascoste. Quindi, per prendere in considerazione le categorie "mancanti" è generalmente necessario usare una statistica percentuale.

Nota: questo controllo non ha alcun effetto sui valori mancanti di sistema che sono sempre esclusi dal grafico.

Statistiche riassuntive e valori dei casi. È possibile selezionare una delle seguenti opzioni per escludere i casi che contengono valori mancanti:

- **Esclusione listwise per ottenere una base uniforme per il grafico.** Se qualsiasi variabile del grafico contiene un caso con un valore mancante, l'intero caso verrà escluso dal grafico.
- **Esclusione di singole variabili per ottimizzare l'uso dei dati.** Se una variabile specifica contiene casi con valori mancanti, tali casi vengono esclusi durante l'analisi della variabile.

Nelle figure seguenti vengono illustrate le differenze tra l'esclusione listwise dei casi mancanti e l'esclusione dei casi mancanti variabile per variabile. Nelle figure è riportato un grafico a barre per ciascuna delle due opzioni.

Figura 46-7
Esclusione listwise dei valori mancanti

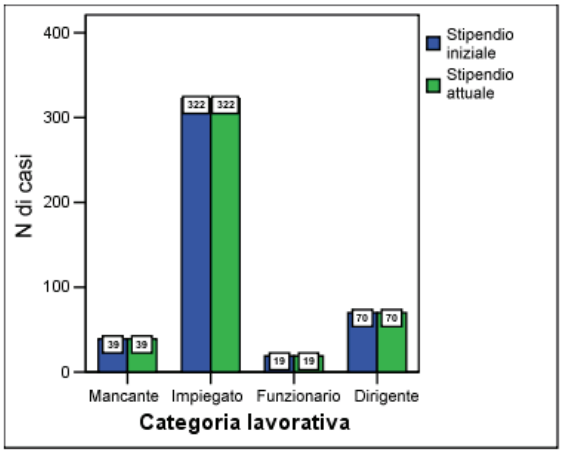
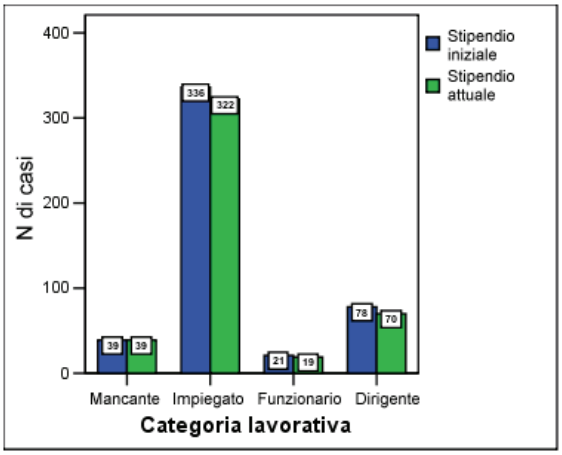


Figura 46-8
Esclusione dei valori mancanti variabile per variabile



I dati comprendono alcuni valori mancanti di sistema nelle variabili per *Stipendio attuale* e *Categoria lavorativa*. In alcuni altri casi, è stato inserito il valore 0 che è stato definito come mancante. In entrambi i grafici è selezionata l'opzione Visualizza i gruppi definiti da valori mancanti, che aggiunge la categoria *Mancanti* alle altre categorie lavorative visualizzate. Nelle etichette

delle barre di ognuno dei grafici sono visualizzati i valori della funzione di rappresentazione *Numero di casi*.

In entrambi i grafici, 26 casi hanno un valore mancante di sistema per la categoria lavorativa e 13 casi hanno il valore mancante definito dall'utente (0). Nel grafico relativo all'esclusione listwise, il numero dei casi è lo stesso per entrambe le variabili in ogni raggruppamento di barre, perché se esiste un valore mancante il caso viene escluso per tutte le variabili. Nel grafico relativo all'esclusione variabile per variabile, il numero di casi non mancanti per ogni variabile viene tracciato indipendentemente dai valori mancanti delle altre variabili.

Modelli

Il modello di grafico permette di applicare gli attributi di un grafico a un altro grafico. Quando si apre un grafico nell'Editor di grafici è possibile salvarlo come modello. Quindi, è possibile applicare il modello specificandolo al momento della creazione o applicandolo in seguito nell'Editor dei grafici.

Modello default. Questo è il modello specificato dalle opzioni. Per accedere alle opzioni, selezionare Opzioni nel menu Modifica dell'Editor dei dati, quindi fare clic sulla scheda Grafici. Il modello predefinito viene applicato per prima, quindi può essere sovrascritto da qualsiasi altro modello.

File modello. Fare clic su Aggiungi per specificare uno o più modelli tramite la finestra di dialogo standard per la selezione dei file. Questi modelli vengono applicati nell'ordine in cui compaiono. Quindi, i modelli alla fine dell'elenco possono avere la priorità rispetto a quelli all'inizio dell'elenco.

Dimensioni dei grafici e riquadri

Dimensioni grafico. Specificare una percentuale maggiore di 100 per ingrandire il grafico o minore di 100 per ridurlo. La percentuale è relativa alla dimensione predefinita del grafico.

Riquadri. In presenza di più colonne di riquadri, selezionare Testo a capo nei riquadri per consentire il ritorno a capo su più righe all'interno dei riquadri invece di adattare il contenuto alla riga. Se questa opzione non è selezionata, i riquadri vengono ridotti per consentirne l'adattamento alla riga.

Capitolo
47

Strumenti

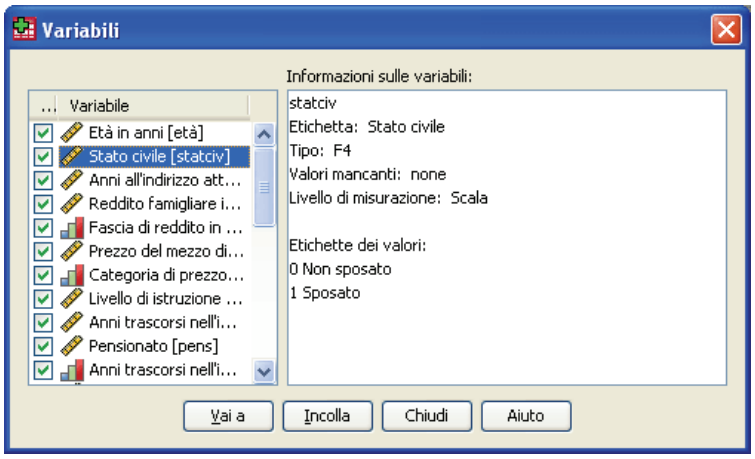
In questo capitolo vengono descritte le funzioni disponibili nel menu Utilità e la possibilità di riordinare gli elenchi delle variabili di destinazione.

Informazioni sulla variabile

Nella finestra di dialogo Variabili vengono visualizzate informazioni sulla definizione della variabile correntemente selezionata. Tali informazioni includono:

- Etichetta di valore
- Formato
- Valori mancanti definiti dall'utente
- Etichette dei valori
- Livello di misurazione

Figura 47-1
Finestra di dialogo Variabili



Visibile. La colonna Visibile nell'elenco delle variabili indica se la variabile è visibile nell'Editor dei dati e negli elenchi di variabili presenti nelle finestre di dialogo. La visibilità viene controllata da insiemi di variabili. Per ulteriori informazioni, vedere Insiemi di variabili a pag. 517.

Vai a. Passa alla variabile selezionata nella finestra dell'Editor dei dati.

Incolla. Incolla le variabili selezionate all'interno della finestra di sintassi designata nel punto in cui è posizionato il cursore.

Per modificare le definizioni delle variabili, utilizzare la Visualizzazione variabili nell'Editor dei dati.

Per ottenere informazioni sulle variabili

- Dai menu, scegliere:
  Strumenti
  Variabili...
- Selezionare la variabile per cui si desidera visualizzare informazioni sulla relativa definizione.

Commenti sul file di dati

È possibile includere commenti descrittivi in un file di dati. Per i file di dati in formato SPSS Statistics, questi commenti vengono salvati con il file di dati.

Per aggiungere, modificare, eliminare o visualizzare i commenti sul file di dati:

- Dai menu, scegliere:
  Strumenti
  Commenti sul file dati...
- Per visualizzare i commenti nel Viewer, selezionare Visualizza i commenti nell'output.

Le dimensioni dei commenti sono illimitate, tuttavia le singole righe dei commenti non possono superare 80 byte, corrispondenti a 80 caratteri nelle lingue a un byte. Alle righe con dimensioni superiori verrà applicato il ritorno a capo dopo l'ottantesimo carattere. I commenti vengono visualizzati con lo stesso tipo di carattere dell'output testuale, in modo da rifletterne la modalità di visualizzazione nel Viewer.

Un indicatore di data (la data corrente tra parentesi) viene automaticamente aggiunto alla fine della lista di commenti ogni volta che si aggiungono o si modificano i commenti. In questo modo, tuttavia, se si modifica un commento esistente o si inserisce un nuovo commento tra commenti esistenti, è possibile che le date associate ai commenti risultino ambigue.

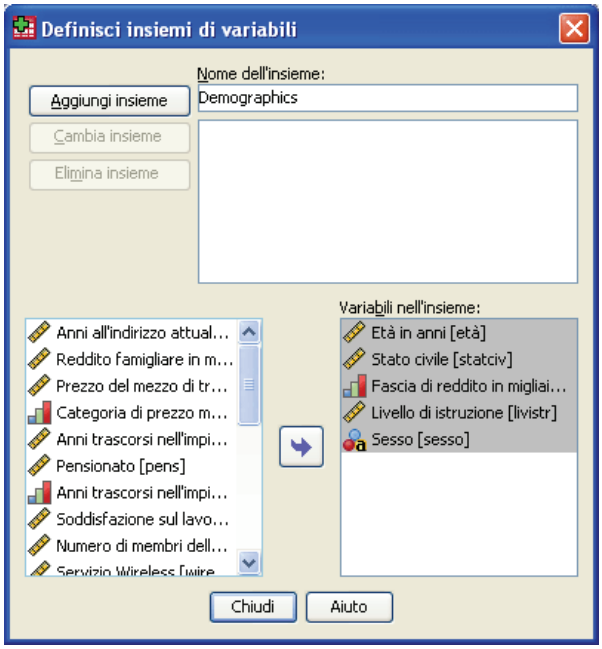
Insiemi di variabili

È possibile limitare il numero di variabili da visualizzare nell'Editor dei dati e negli elenchi di variabili presenti nelle finestre di dialogo mediante la definizione e l'uso di insiemi di variabili. Ciò risulta particolarmente utile per i file di dati che contengono un numero elevato di variabili. L'utilizzo di piccoli insiemi di variabili semplifica la ricerca e la selezione delle variabili per l'analisi.

Definisci insiemi di variabili

In questa finestra di dialogo è possibile creare sottoinsiemi di variabili da visualizzare nell'Editor dei dati e negli elenchi di variabili presenti nelle finestre di dialogo. Gli insiemi di variabili definiti vengono salvati in file di dati in formato SPSS Statistics.

Figura 47-2
Finestra di dialogo Definisci insiemi di variabili



Nome dell'insieme. I nomi degli insiemi possono essere lunghi fino a 64 byte. È possibile utilizzare qualsiasi carattere, inclusi gli spazi.

Variabili nell'insieme. In un insieme è possibile inserire qualsiasi combinazione di variabili numeriche e stringa. L'ordine delle variabili nell'insieme non influenza in alcun modo l'ordine di visualizzazione delle variabili nell'Editor dei dati o negli elenchi di variabili presenti nelle finestre di dialogo. Una variabile può appartenere a più insiemi di variabili.

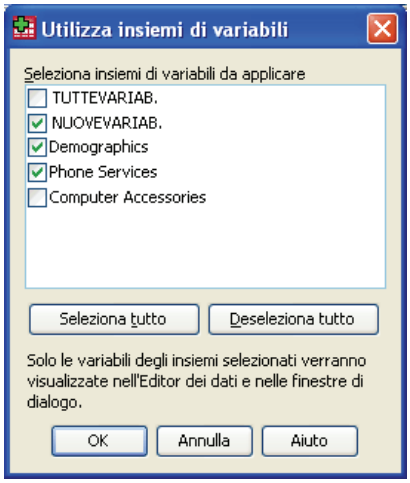
Per definire insiemi di variabili

- Dai menu, scegliere:
  Strumenti
  Definisci insiemi di variabili...
- Selezionare le variabili che si desidera includere nell'insieme.
- Immettere un nome per l'insieme (fino a 64 byte).
- Fare clic su Aggiungi insieme.

Utilizza insiemi di variabili

Utilizza set di variabili limita le variabili visualizzate nell'Editor dei dati e negli elenchi di variabili presenti nelle finestre di dialogo al numero di variabili negli insiemi selezionati.

Figura 47-3
Finestra di dialogo Utilizza set di variabili



- Gli insiemi di variabili visualizzati nell'Editor dei dati e negli elenchi di variabili presenti nelle finestre di dialogo sono l'unione di tutti gli insiemi selezionati.
- Una variabile può essere inclusa in più insiemi selezionati.
- L'ordine delle variabili negli insiemi selezionati e l'ordine degli insiemi selezionati non influenza l'ordine di visualizzazione delle variabili nell'Editor dei dati o negli elenchi di variabili presenti nelle finestre di dialogo.
- Sebbene gli insiemi di variabili definiti vengano salvati in file di dati in formato SPSS Statistics, l'elenco degli insiemi correntemente selezionati viene ripristinato sul valore predefinito, ovvero gli insiemi di sistema, ogni volta che si apre il file di dati.

L'elenco degli insiemi di variabili disponibili include tutti gli insiemi definiti per l'insieme di dati attivo, più due insiemi di sistema:

- **TUTTEVARIAB.** Questo insieme contiene tutte le variabili nel file di dati, incluse le nuove variabili create durante una sessione.
- **NUOVEVARIAB.** Questo insieme contiene solo le nuove variabili create durante la sessione.

  Nota: anche se si salva il file di dati dopo la creazione di nuove variabili, le nuove variabili rimangono incluse nell'insieme *NUOVEVARIAB* fin quando il file di dati non viene chiuso e riaperto.

Almeno un insieme di variabili deve essere selezionato. Se l'opzione *TUTTEVARIAB* è selezionata, tutti gli altri insiemi selezionati non saranno visibili, dal momento che questo insieme contiene tutte le variabili.

Per selezionare gli insiemi di variabili da visualizzare

▶ Dai menu, scegliere:

Strumenti
  Utilizza insiemi di variabili...

▶ Selezionare gli insiemi di variabili definiti che includono le variabili che si desidera visualizzare nell'Editor dei dati e negli elenchi di variabili presenti nelle finestre di dialogo.

Per visualizzare tutte le variabili

▶ Dai menu, scegliere:
Strumenti
Mostra tutte le variabili

Riordinamento degli elenchi di variabili di destinazione

Le variabili compaiono negli elenchi di destinazione della finestra di dialogo nello stesso ordine in cui vengono selezionate negli elenchi sorgente. Se si desidera cambiare l'ordine delle variabili in un elenco di destinazione, ma non si vuole deselezionarle tutte e riselezionarle in un nuovo ordine, è possibile spostare le variabili in alto o in basso nell'elenco di destinazione utilizzando il tasto Ctrl (Macintosh: tasto Comando) insieme ai tasti freccia su e freccia giù. È possibile spostare simultaneamente più variabili adiacenti (raggruppate insieme). Non è possibile spostare gruppi di variabili non adiacenti.

Capitolo
48

Opzioni

Nella finestra di dialogo Opzioni è possibile impostare numerosi elementi, ad esempio:

- Il giornale di sessione, che consente di registrare tutti i comandi eseguiti in ciascuna sessione.
- L'ordine di visualizzazione delle variabili negli elenchi sorgente delle finestre di dialogo.
- Gli elementi visualizzati e nascosti nei risultati del nuovo output.
- Modello di tabelle per nuove tabelle pivot.
- I formati di valuta personalizzati.

Per modificare le impostazioni delle opzioni

▶ Dai menu, scegliere:
Modifica
Opzioni...

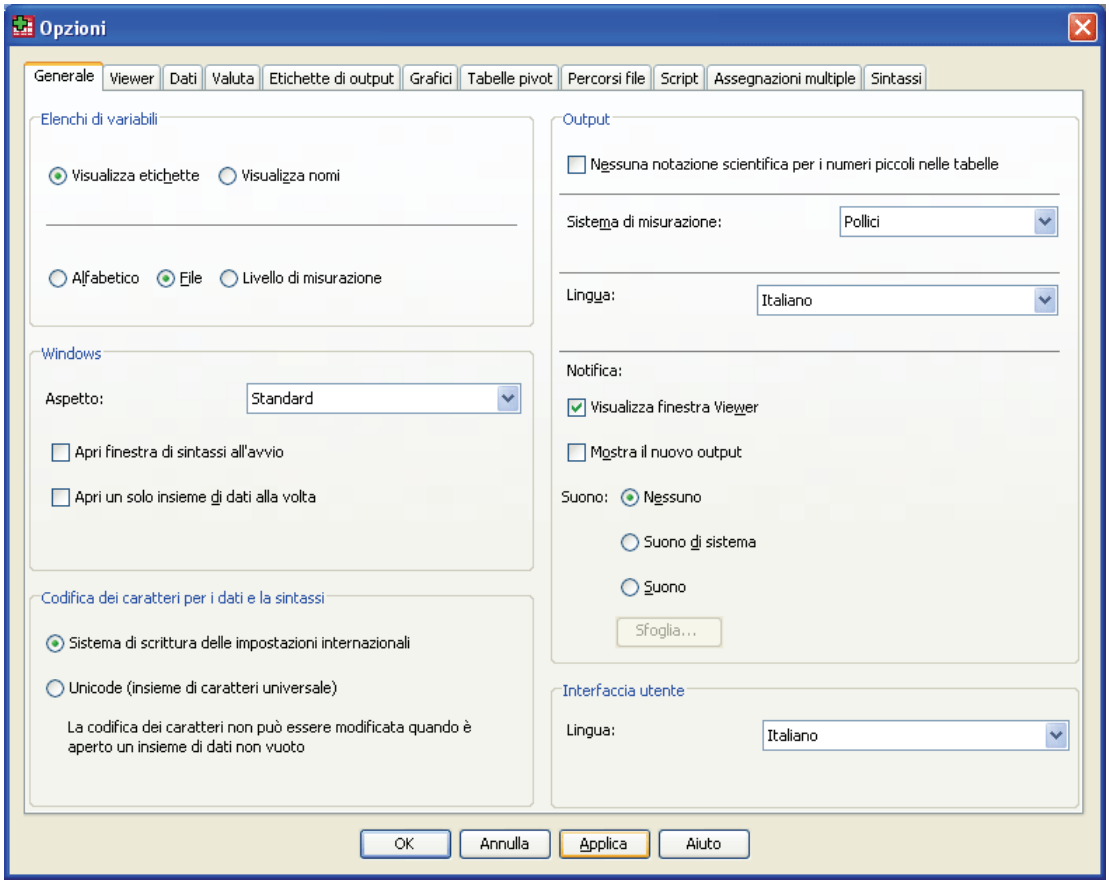
▶ Fare clic sulle schede corrispondenti alle impostazioni che si desidera modificare.

▶ Modificare le impostazioni.

▶ Fare clic su OK o su Applica.

Opzioni generali

Figura 48-1
Finestra di dialogo Opzioni: scheda Generale



Elenchi di variabili

Queste impostazioni consentono di controllare la visualizzazione delle variabili negli elenchi delle finestre di dialogo. È possibile visualizzare i nomi o le etichette delle variabili. I nomi o le etichette possono essere visualizzati in ordine alfabetico, in ordine di file oppure raggruppati per livello di misurazione. L'ordine di visualizzazione viene applicato solo agli elenchi di variabili sorgente. Gli elenchi di variabili di destinazione riflettono sempre l'ordine di selezione delle variabili.

Windows

Aspetto. Consente di controllare l'aspetto di base delle finestre e delle finestre di dialogo.

Apri finestra di sintassi all'avvio. Le finestre di sintassi sono finestre di file di testo in cui è possibile immettere, modificare o eseguire i comandi. Se si utilizza spesso la sintassi dei comandi, selezionare questa opzione per aprire automaticamente una finestra di sintassi all'avvio di ogni sessione. Ciò risulta utile soprattutto per gli utenti esperti che preferiscono utilizzare la sintassi dei comandi anziché le finestre di dialogo.

Apri un solo insieme di dati alla volta. Consente di chiudere la sorgente dati aperta ogni volta che ne viene aperta una diversa mediante i menu e le finestre di dialogo. Per impostazione predefinita, ogni volta che vengono utilizzati i menu e le finestre di dialogo per aprire una nuova sorgente dati, tale sorgente dati viene aperta in una nuova finestra dell'Editor dei dati e qualsiasi altra sorgente dati aperta in altre finestre dell'Editor dei dati rimane aperta e disponibile durante la sessione finché non viene esplicitamente chiusa.

La selezione di questa opzione ha effetto immediato ma non determina la chiusura di altri insiemi di dati che erano aperti nel momento in cui è stata modificata l'impostazione. Questa impostazione non ha alcun effetto sulle sorgenti dati aperte mediante la sintassi dei comandi che si basa sui comandi DATASET per il controllo di più insiemi di dati.Per ulteriori informazioni, vedere Utilizzo di più sorgenti dati in Capitolo 6 a pag. 102.

Codifica dei caratteri per i dati e la sintassi

Controlla il comportamento predefinito che determina la codifica per la lettura e la scrittura di file di dati e di sintassi. È possibile modificare questa impostazione solo quando non vi sono sorgenti dati aperte; l'impostazione rimane attiva per le sessioni successive finché non viene modificata in modo esplicito.

- **Sistema di scrittura delle impostazioni internazionali.** Consente di utilizzare le impostazioni internazionali correnti per determinare la codifica per la lettura e la scrittura dei file. Ciò viene definito **modalità di pagina di codice**.
- **Unicode (insieme di caratteri universale).** Consente di utilizzare la codifica Unicode (UTF-8) per la lettura e la scrittura dei file. Ciò viene definito **modalità Unicode**.

Vi sono una serie di implicazioni importanti riguardo la modalità e i file Unicode:

- I file di dati in formato SPSS Statisticse i file di sintassi salvati nella codifica Unicode non dovrebbero essere utilizzati nelle versioni di SPSS Statistics precedenti alla 16.0. Per i file di sintassi, è possibile specificare la codifica al momento del salvataggio del file. I file di dati possono essere aperti in modalità di pagina di codice e risalvati nel caso in cui si desideri leggerli con le versioni precedenti.
- Quando i file di dati che sono in modalità di pagina di codice vengono letti in modalità Unicode, la larghezza definita di tutte le variabili di stringa viene triplicata. Per impostare automaticamente la larghezza di ogni variabile di stringa sul valore più lungo osservato per tale variabile, selezionare Minimizza le larghezze delle stringhe sulla base dei valori osservati.

Output

Nessuna notazione scientifica per i numeri piccoli nelle tabelle. Consente di sopprimere la visualizzazione della notazione scientifica per valori decimali piccoli nell'output. I valori decimali molto piccoli verranno visualizzati come 0 (oppure 0,000).

Sistema di misurazione. Il sistema di misura utilizzato (punti, pollici o centimetri) per specificare attributi quali, ad esempio, i margini delle celle delle tabelle pivot, la larghezza delle celle e lo spazio tra le tabelle per la stampa.

Lingua. Consente di controllare la lingua utilizzata nell'output. Non è disponibile per l'output testuale semplice. L'elenco di lingue disponibili dipende dai file di lingua correntemente installati. (Nota: *non* incide sulla lingua dell'interfaccia utente).

Nota: gli script personalizzati che si basano su stringhe di testo specifiche per la lingua nell'output potrebbero non essere eseguiti correttamente quando viene modificata la lingua dell'output. Per ulteriori informazioni, vedere Opzioni: Script a pag. 537.

Notifica. Consente di impostare le modalità di notifica automatica del completamento dell'esecuzione di una procedura e della disponibilità dei risultati nel Viewer.

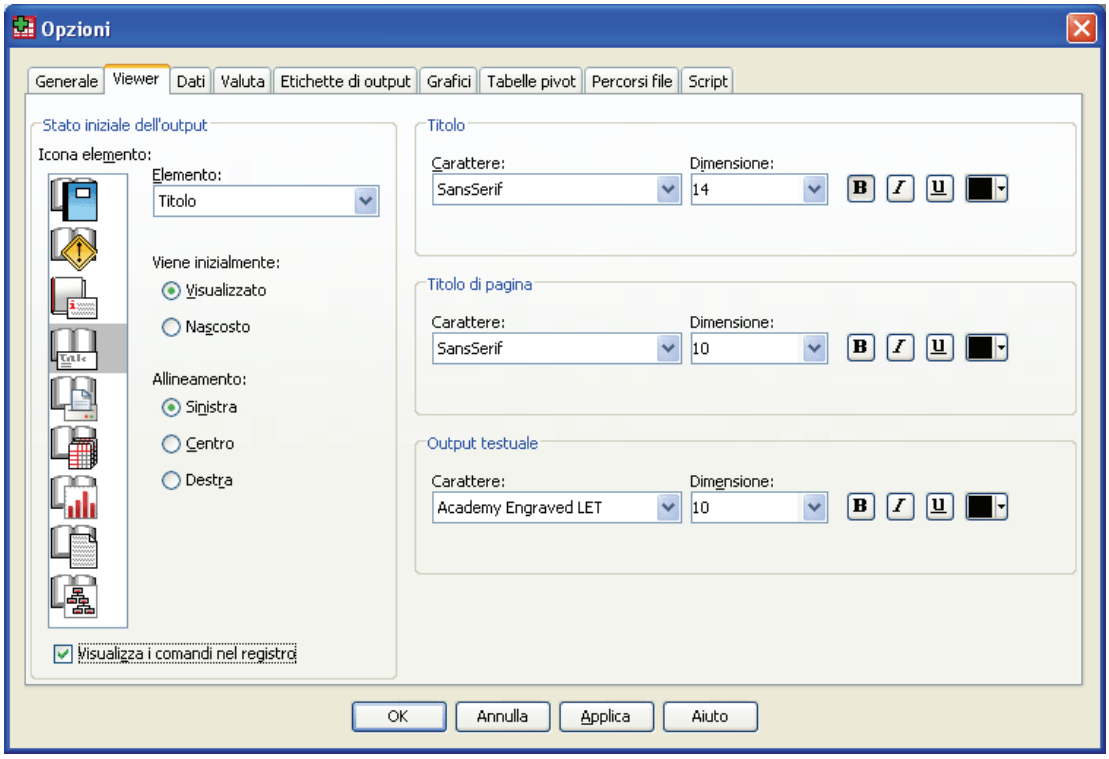
Interfaccia utente

Controlla la lingua utilizzata nei menu, nelle finestre di dialogo e in altre funzioni dell'interfaccia utente. (Nota: *non* incide sulla lingua dell'output).

Opzioni Viewer

Le opzioni di visualizzazione dell'output del Viewer influiscono solo sul nuovo output prodotto dopo la modifica delle impostazioni. Le modifiche apportate a tali impostazioni non influiscono sull'output già visualizzato nel Viewer.

Figura 48-2
Finestra di dialogo Opzioni: Scheda Viewer



Stato iniziale dell'output. Consente di impostare gli elementi che verranno visualizzati o nascosti automaticamente ad ogni esecuzione di una procedura e le modalità di allineamento iniziale di tali elementi. È possibile impostare la visualizzazione dei seguenti elementi: registro, avvisi, note, titoli, tabelle pivot, grafici, diagrammi ad albero e output testuale. È inoltre possibile attivare o disattivare la visualizzazione dei comandi del registro. È possibile copiare la sintassi dei comandi dal registro e salvarla in un file di sintassi.

Nota: Nel Viewer tutti gli elementi dell'output vengono allineati a sinistra. Le impostazioni di giustificazione vengono applicate solo all'output stampato. Gli elementi centrati e allineati a destra sono contrassegnati da un piccolo simbolo situato sopra e a sinistra di ciascun elemento.

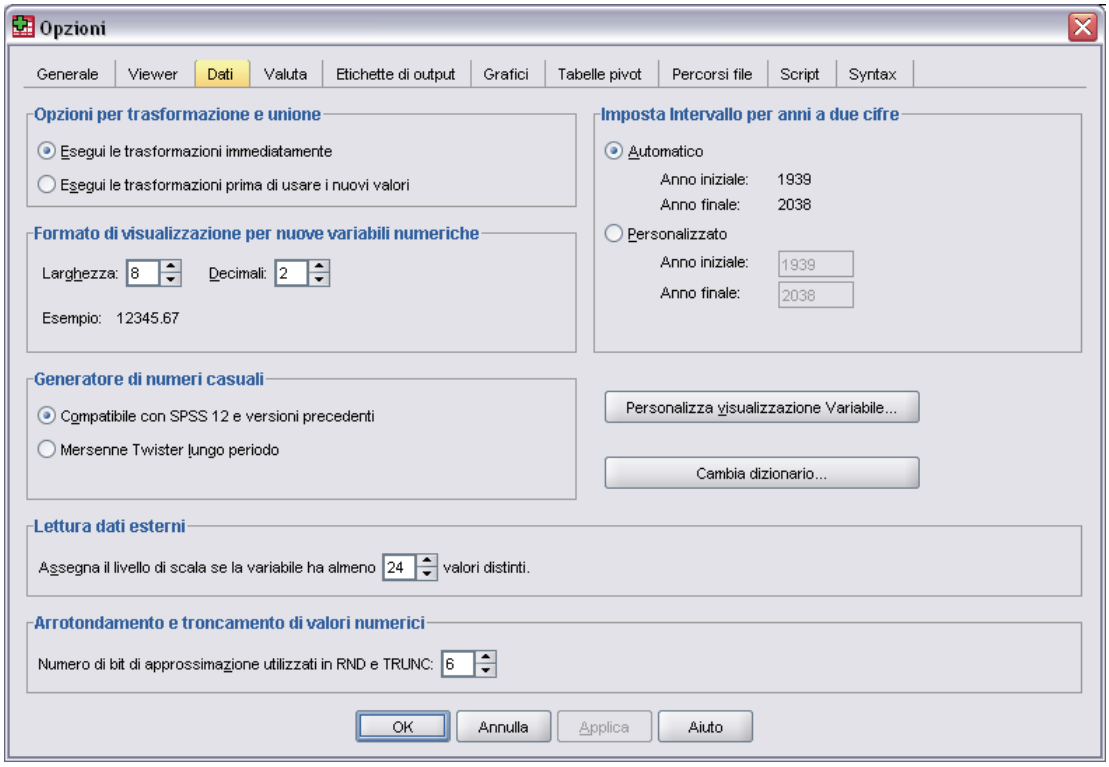
Titolo. Consente di impostare lo stile, la dimensione e il colore del carattere per i titoli del nuovo output.

Titolo di pagina. Consente di impostare lo stile, la dimensione e il colore del carattere per i titoli della nuova pagina e per i titoli della pagina generati dalla sintassi dei comandi `TITLE` e `SUBTITLE` o creati tramite l'opzione Nuovo titolo pagina nel menu Inserisci.

Output testuale. Carattere utilizzato per l'output testuale. Per l'output testuale è necessario utilizzare un carattere proporzionale (a spaziatura fissa). Se si seleziona un carattere proporzionale, l'output non risulterà allineato correttamente nelle tabelle.

Opzioni: Dati

Figura 48-3
Finestra di dialogo Opzioni: Scheda Dati



Opzioni per trasformazione e unione. Ogni volta che viene eseguito un comando, viene letto il file dati. Per alcune trasformazioni di dati (ad esempio Calcola e Ricodifica) e di file (ad esempio Aggiungi variabili e Aggiungi casi) non è necessario il passaggio separato dei dati e l'esecuzione di questi comandi può essere rimandata finché non vengono letti i dati per l'esecuzione di un altro comando, ad esempio una procedura statistica o di creazione di grafico.

- Per file dati di grandi dimensioni, in cui la lettura dei dati rischia di impiegare un certi tempo, selezionare Esegui le trasformazioni prima di usare i nuovi valori per rimandare l'esecuzione e risparmiare tempo di elaborazione. Quando questa opzione è selezionata, i risultati delle trasformazioni effettuate tramite finestre di dialogo come Calcola variabile non appaiono immediatamente nell'Editor dei dati; le nuove variabili vengono visualizzate senza i valori dei dati; e i valori dei dati nell'Editor dei dati non possono essere modificati in presenza di trasformazioni in sospeso. I comandi che leggono dati, come le procedure statistiche o di creazione di grafico, eseguono le trasformazioni in sospeso e aggiornano i dati visualizzati nell'Editor dei dati. In alternativa, è possibile usare Esegui trasformazioni in sospeso nel menu Trasforma.
- Con l'impostazione predefinita di Esegui le trasformazioni immediatamente, dopo avere incollato una sintassi di comando dalle finestre di dialogo, viene incollato un comando `EXECUTE` dopo ogni comando di trasformazione. Per ulteriori informazioni, vedere Comandi EXECUTE multipli in Capitolo 13 a pag. 281.

Formato di visualizzazione per nuove variabili numeriche. Consente di modificare la larghezza di visualizzazione predefinita e il numero di decimali delle nuove variabili numeriche. Per le nuove variabili stringa non è previsto un formato di visualizzazione predefinito. Se un valore è troppo grande per il formato di visualizzazione specificato, verranno innanzitutto arrotondate le cifre decimali e quindi i valori verranno convertiti in notazione scientifica. I formati di visualizzazione non producono alcun effetto sui valori interni. Ad esempio, il valore 123456,78 può essere arrotondato a 123457 per la visualizzazione, ma nei calcoli verrà utilizzato il valore originale non arrotondato.

Imposta intervallo per anni a due cifre. Consente di definire l'intervallo di anni per le variabili in formato data immesse e/o visualizzate con un anno a due cifre (per esempio, 28/10/86, 29-OTT-87). L'impostazione automatica dell'intervallo è basata sull'anno corrente, con inizio 69 anni prima e con termine 30 anni dopo l'anno corrente (aggiungendo l'anno corrente si ottiene un intervallo totale di 100 anni). Per un intervallo personalizzato, l'anno finale viene determinato automaticamente in base al valore immesso come anno iniziale.

Generatore di numeri casuale. Sono disponibili due diversi generatori di numeri casuali:

- **Compatibile con la versione 12.** Il generatore di numeri casuali utilizzato nella versione 12 e precedenti. Utilizzare questo generatore se è necessario riprodurre i risultati randomizzati generati in versioni precedenti in base a un valore di seme specificato.
- **Mersenne Twister.** Un generatore di numeri casuali più affidabile a scopi di simulazione. Se la riproduzione di risultati randomizzati dalla versione 12 o precedenti non rappresenta un problema, utilizzare questo generatore.

Lettura dati esterni. Per i dati letti da formati di file esterni e da file di dati SPSS Statistics precedenti (antecedenti alla versione 8.0), è possibile specificare il numero minimo di valori di dati per la variabile numerica usata per classificare la variabile come scala o nominale. Le variabili con un numero di valori distinti inferiore a quello specificato vengono classificate come nominali.

Arrotondamento e troncamento di valori numerici. Per le funzioni `RND` e `TRUNC`, questa impostazione controlla la soglia predefinita per l'arrotondamento per eccesso dei valori molto vicini a un limite di arrotondamento. L'impostazione è espressa come numero di bit ed è impostata su 6 al momento dell'installazione, un livello che dovrebbe essere sufficiente per la maggior parte delle applicazioni. L'impostazione del numero di bit su 0 produce gli stessi risultati prodotti nella versione 10. L'impostazione del numero di bit su 10 produce gli stessi risultati prodotti nelle versioni 11 e 12.

- Per la funzione `RND`, questa impostazione specifica la differenza massima (per difetto) di bit meno significativi tra il valore da arrotondare e la soglia di arrotondamento per eccesso che consente comunque l'arrotondamento. Ad esempio, quando un valore compreso tra 1,0 e 2,0 viene arrotondato al numero intero più vicino, questa impostazione indica di quanto il valore può essere inferiore a 1,5 (soglia prevista per l'arrotondamento a 2,0) per essere comunque arrotondato a 2,0.
- Per la funzione `TRUNC`, questa impostazione specifica la differenza massima (per difetto) di bit meno significativi tra il valore da troncare e la soglia di arrotondamento più vicina che consente comunque l'arrotondamento per eccesso prima del troncamento. Ad esempio, quando un valore compreso tra 1,0 e 2,0 viene troncato al numero intero più vicino, questa impostazione indica di quanto il valore può essere inferiore a 2,0 per poter essere comunque arrotondato a 2,0.

Personalizza la visualizzazione Variabile. Controlla la visualizzazione e l'ordine predefiniti degli attributi nella Visualizzazione variabili. Per ulteriori informazioni, vedere Modifica della Visualizzazione variabili predefinita a pag. 527.

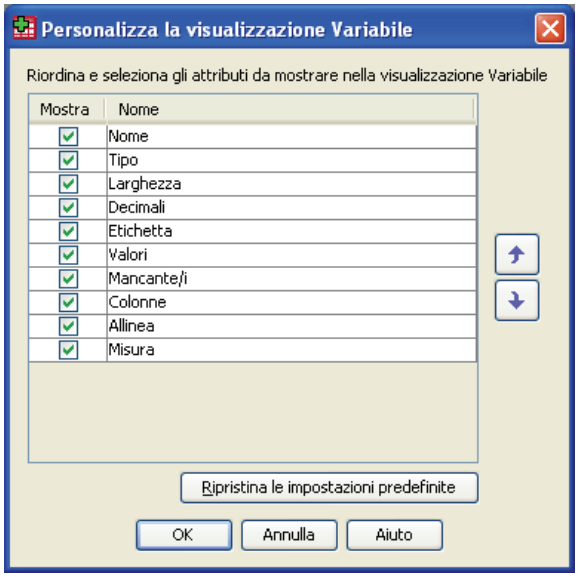
Cambia dizionario. Controlla la versione della lingua del dizionario utilizzata per il controllo dell'ortografia delle voci nella Visualizzazione variabili. Per ulteriori informazioni, vedere Controllo ortografico in Capitolo 5 a pag. 92.

Modifica della Visualizzazione variabili predefinita

È possibile utilizzare Personalizza la visualizzazione Variabile per controllare quali attributi vengono mostrati per impostazione predefinita nella Visualizzazione variabili (ad esempio nome, tipo, etichetta) e l'ordine in cui vengono mostrati.

Fare clic su Personalizza la visualizzazione Variabile.

Figura 48-4
Personalizza la visualizzazione Variabile (impostazione predefinita)



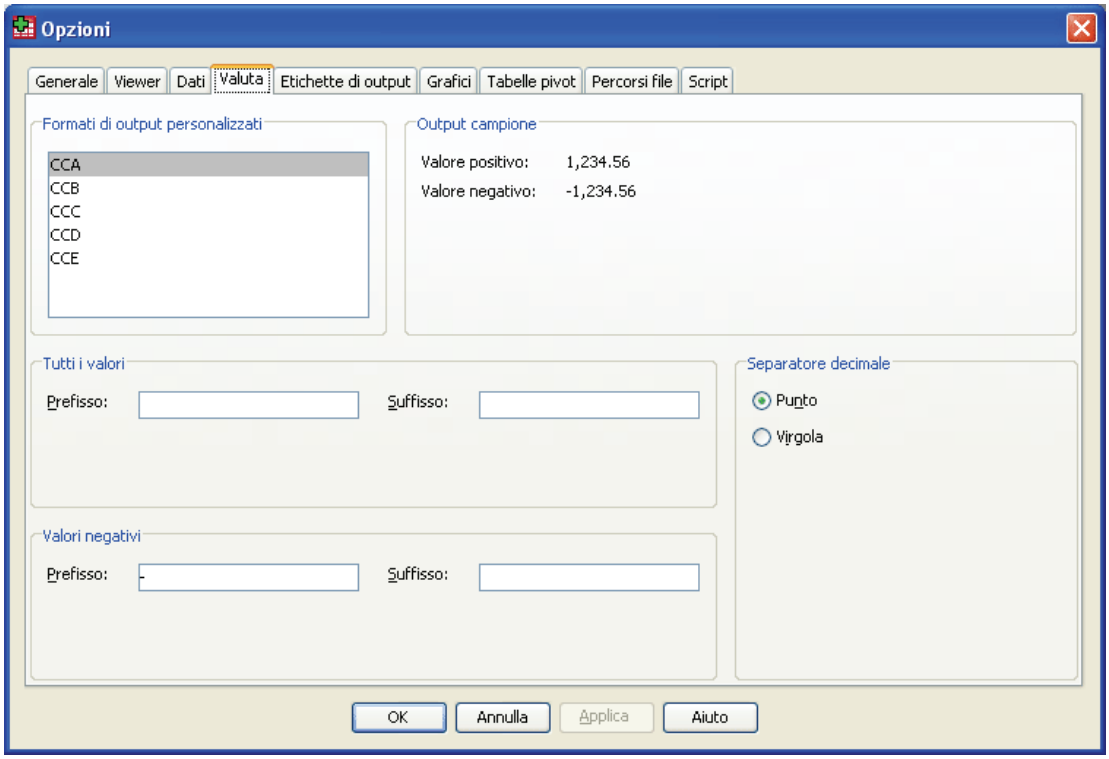
- Selezionare gli attributi di variabile da visualizzare.
- Utilizzare i pulsanti di spostamento verso l'alto e verso il basso per cambiare l'ordine di visualizzazione degli attributi.

Opzioni: Valuta

È possibile creare al massimo cinque formati di visualizzazione della valuta personalizzati che possono includere caratteri speciali di prefisso o suffisso e un particolare trattamento dei valori negativi.

I nomi dei cinque formati di valuta personalizzati sono CCA, CCB, CCC, CCD e CCE. Non è possibile modificare i nomi dei formati o aggiungerne altri. Per modificare un formato di valuta personalizzato, selezionarne il nome nell'elenco di origine e apportare le modifiche desiderate.

Figura 48-5
Finestra di dialogo Opzioni: Scheda Valuta



I prefissi, i suffissi e i separatori decimali definiti per i formati di valuta personalizzati vengono utilizzati solo ai fini della visualizzazione. Non è possibile inserire valori nell'Editor dei dati utilizzando i caratteri di valuta personalizzati.

Per creare formati di valuta personalizzati

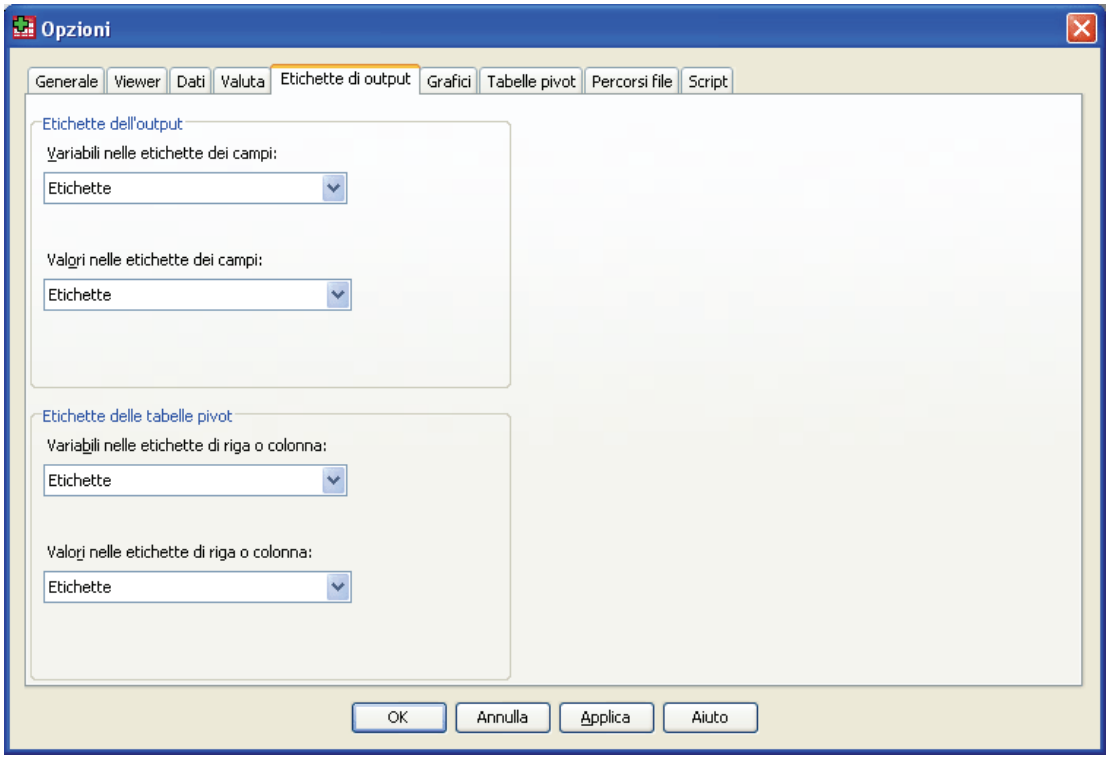
- Fare clic sulla scheda Valuta.
- Selezionare uno dei formati valuta dall'elenco (CCA, CCB, CCC, CCD e CCE).
- Specificare il prefisso, il suffisso e il separatore decimale.
- Fare clic su OK o su Applica.

Opzioni: Etichette di output

La scheda Etichette di output consente di impostare la visualizzazione delle informazioni sulle variabili e sui valori dei dati nella vista riassuntiva e nelle tabelle pivot. È possibile visualizzare i nomi delle variabili, le etichette e i valori effettivi delle variabili definite, le etichette dei valori definite o una combinazione.

Le etichette dei valori e delle variabili descrittive (Visualizzazione variabili nell'Editor dei dati, colonne *Etichetta* e *Valori*) spesso agevolano l'interpretazione dei risultati. Le etichette molto lunghe, tuttavia, talvolta possono creare dei problemi.

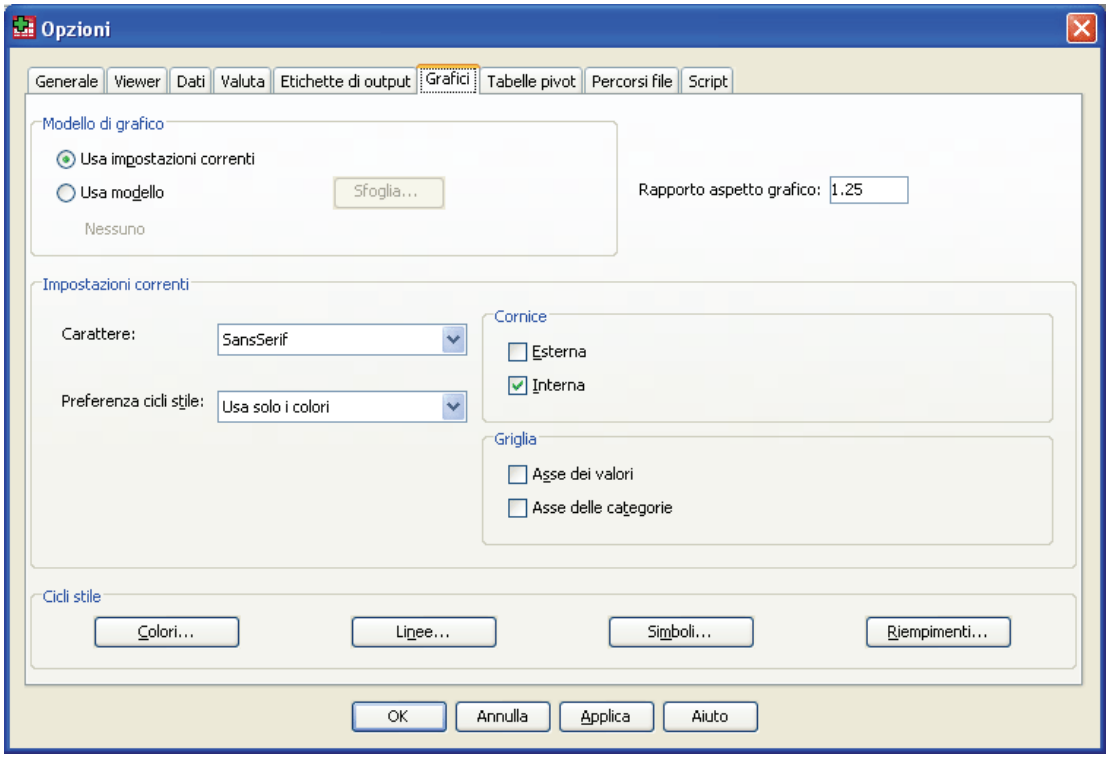
Figura 48-6
Finestra di dialogo Opzioni: Scheda Etichette di output



Le opzioni delle etichette di output vengono applicate solo al nuovo output prodotto in seguito alla modifica delle impostazioni. Le modifiche apportate a tali impostazioni non influiscono sull'output già visualizzato nel Viewer. Queste impostazioni vengono applicate solo all'output delle tabelle pivot. L'output testuale rimarrà invariato.

Grafici: opzioni

Figura 48-7
Finestra di dialogo Opzioni: Scheda Grafici



Modello di grafico. Per i nuovi grafici è possibile utilizzare le impostazioni selezionate in questo gruppo oppure le impostazioni di un file del modello di grafico. Fare clic su Sfoglia per selezionare un modello di grafico. Per creare un modello di grafico, è necessario creare un file con gli attributi desiderati e quindi salvarlo come modello (menu File, Salva modello).

Proporzione del grafico. Il rapporto tra altezza e larghezza della cornice esterna dei nuovi grafici. È possibile specificare un rapporto larghezza/altezza compreso tra 0,1 e 10,0. Con valori inferiori a 1, si ottengono grafici con un'altezza maggiore della larghezza. Specificando valori maggiori di 1 si otterranno grafici più larghi che alti. Specificando 1 si otterrà un grafico quadrato. Dopo aver creato un grafico non è possibile modificarne le proporzioni.

Impostazioni correnti. Sono disponibili le seguenti impostazioni:

- **Carattere.** Il carattere utilizzato per tutto il testo dei nuovi grafici.
- **Preferenza ciclo stile.** L'assegnazione iniziale di colori e/o motivi per i nuovi grafici. Usa solo i colori consente di utilizzare solo i colori per differenziare gli elementi del grafico senza usare i motivi. Usa i motivi consente di utilizzare solo gli stili di linea, i simboli o i motivi per differenziare gli elementi del grafico senza usare i colori.
- **Cornice.** Consente di impostare la visualizzazione di cornici interne ed esterne nei nuovi grafici.
- **Griglia.** Consente di impostare la visualizzazione della griglia dell'asse dei valori e delle categorie nei nuovi grafici.

Cicli stile. Consente di personalizzare i colori, gli stili della linea, i simboli e i motivi nei nuovi grafici. È possibile modificare l'ordine dei colori e dei motivi utilizzati al momento della creazione di un nuovo grafico.

Colori degli elementi dei dati

È possibile specificare l'ordine in cui si desidera che vengano utilizzati i colori per gli elementi dei dati, ad esempio le barre e i simboli, nel nuovo grafico. I colori vengono utilizzati ogni volta che si seleziona un'opzione che includa il *colore* nel gruppo Preferenza ciclo stile della finestra di dialogo Grafici: Opzioni principale.

Se, ad esempio, si crea un grafico a barre raggruppato con due gruppi e si seleziona Usa i colori nella finestra di dialogo Grafici: Opzioni principale, i primi due colori nell'elenco Grafici raggruppati verranno utilizzati come colori delle barre nel nuovo grafico.

Per modificare l'ordine di utilizzo dei colori:

- Selezionare Grafici semplici e quindi selezionare un colore utilizzato per i grafici senza categorie.
- Selezionare Grafici raggruppati per modificare il ciclo dei colori per i grafici con le categorie. Per modificare il colore di una categoria, selezionare una categoria e quindi un colore dalla tavolozza.

Se lo si desidera, è possibile:

- Inserire una nuova categoria al di sopra di quella selezionata.
- Spostare una categoria selezionata.
- Eliminare una categoria selezionata.
- Ripristinare la sequenza predefinita.
- Modificare un colore selezionandolo e quindi scegliendo Modifica.

Linee degli elementi dei dati

È possibile specificare l'ordine in cui devono essere utilizzati gli stili per i dati lineari nel grafico. Gli stili delle linee vengono utilizzati ogni volta che nel grafico sono presenti dati lineari e che si seleziona un'opzione che includa i *motivi* nel gruppo Preferenza ciclo stile della finestra di dialogo Grafici: Opzioni principale.

Se, ad esempio, si crea un grafico lineare con due gruppi e si seleziona Usa solo i motivi nella finestra di dialogo Grafici: Opzioni principale, i primi due stili nell'elenco Grafici raggruppati verranno utilizzati come motivi nel nuovo grafico.

Per modificare l'ordine di utilizzo degli stili delle linee:

- Selezionare Grafici semplici e quindi selezionare uno stile di linea utilizzato per i grafici senza categorie.
- Selezionare Grafici raggruppati per modificare il ciclo dei motivi per i grafici lineari con le categorie. Per modificare lo stile della linea di una categoria, selezionare una categoria e quindi uno stile dalla tavolozza.

Se lo si desidera, è possibile:

- Inserire una nuova categoria al di sopra di quella selezionata.
- Spostare una categoria selezionata.
- Eliminare una categoria selezionata.
- Ripristinare la sequenza predefinita.

Simboli degli elementi dei dati

È possibile specificare l'ordine in cui devono essere utilizzati i simboli per gli elementi dei dati nel grafico. Gli stili dei simboli vengono utilizzati ogni volta che nel grafico sono presenti simboli e che si seleziona un'opzione che includa i *motivi* nel gruppo Preferenza ciclo stile della finestra di dialogo Grafici: Opzioni principale.

Se, ad esempio, si crea un grafico a dispersione con due gruppi e si seleziona Usa solo i motivi nella finestra di dialogo Grafici: Opzioni principale, i primi due simboli nell'elenco Grafici raggruppati verranno utilizzati come simboli nel nuovo grafico.

Per modificare l'ordine di utilizzo degli stili dei simboli:

▶ Selezionare Grafici semplici e quindi selezionare un simbolo per i grafici senza categorie.

▶ Selezionare Grafici raggruppati per modificare il ciclo dei motivi per i grafici con le categorie. Per modificare il simbolo di una categoria, selezionare una categoria e quindi un simbolo dalla tavolozza.

Se lo si desidera, è possibile:

- Inserire una nuova categoria al di sopra di quella selezionata.
- Spostare una categoria selezionata.
- Eliminare una categoria selezionata.
- Ripristinare la sequenza predefinita.

Riempimenti degli elementi dei dati

È possibile specificare l'ordine in cui devono essere utilizzati gli stili di riempimento per le aree e le barre nel grafico. Gli stili di riempimento vengono utilizzati ogni volta che nel grafico sono presenti barre o aree e che si seleziona un'opzione che includa i *motivi* nel gruppo Preferenza ciclo stile della finestra di dialogo Grafici: Opzioni principale.

Se, ad esempio, si crea un grafico a barre raggruppato con due gruppi e si seleziona Usa solo i motivi nella finestra di dialogo Grafici: Opzioni principale, i primi due stili nell'elenco Grafici raggruppati verranno utilizzati come motivi di riempimento nel nuovo grafico.

Per modificare l'ordine di utilizzo degli stili di riempimento:

▶ Selezionare Grafici semplici e quindi selezionare un motivo utilizzato per i grafici senza categorie.

- Selezionare Grafici raggruppati per modificare il ciclo dei motivi per i grafici con le categorie. Per modificare il motivo di una categoria, selezionare una categoria e quindi un motivo dalla tavolozza.

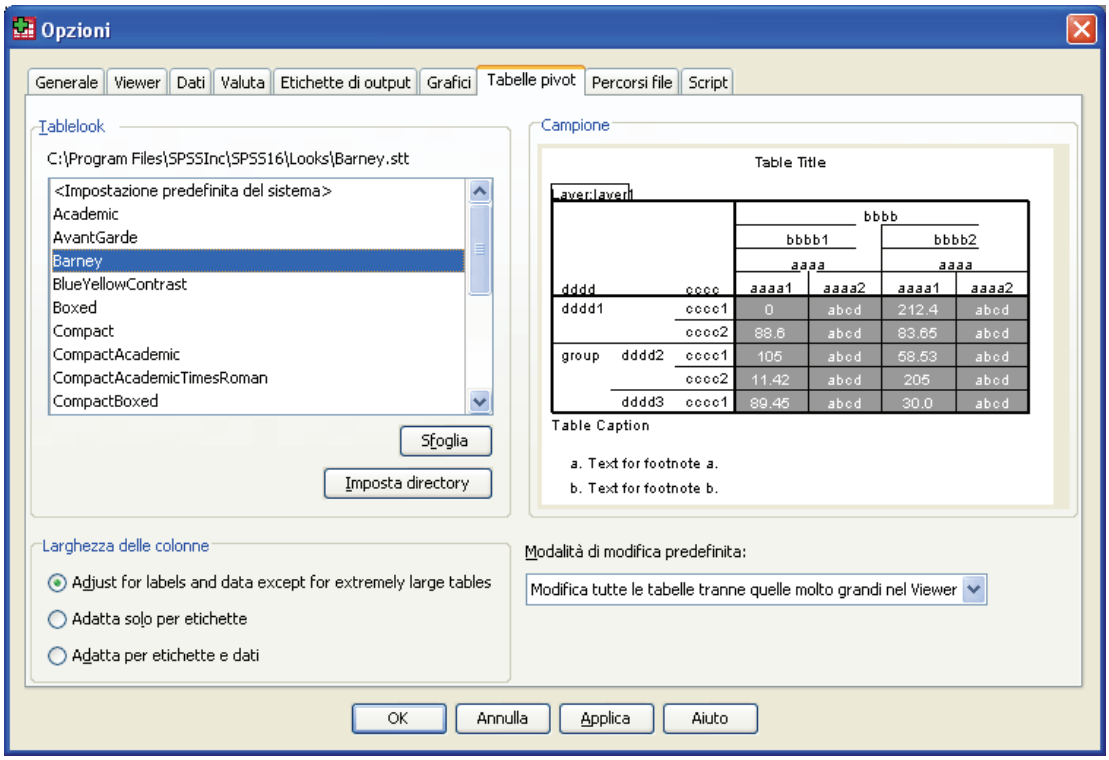
Se lo si desidera, è possibile:

- Inserire una nuova categoria al di sopra di quella selezionata.
- Spostare una categoria selezionata.
- Eliminare una categoria selezionata.
- Ripristinare la sequenza predefinita.

Opzioni tabella pivot

Le opzioni della tabella pivot impostano il TableLook predefinito utilizzato per l'output delle nuove tabelle pivot. Utilizzando i TableLook è possibile impostare numerosi attributi delle tabelle pivot, ad esempio la visualizzazione e la larghezza della griglia, lo stile, la dimensione e il colore del carattere e i colori di sfondo.

Figura 48-8
Finestra di dialogo Opzioni: Scheda Tabelle pivot



TableLook. Selezionare un TableLook dall'elenco dei file e quindi fare clic su OK o su Applica. È possibile utilizzare uno dei modelli di tabelle disponibili in SPSS Statistics oppure crearne uno personalizzato nell'Editor di tabelle pivot (scegliere Modelli di tabelle dal menu Formato).

- **Sfoglia.** Consente di selezionare un TableLook da un'altra directory.
- **Imposta directory.** Consente di modificare la directory predefinita dei TableLook.

Nota: i tablelook creati nelle versioni precedenti di SPSS Statistics non possono essere utilizzati nella versione 16.0 o nelle versioni successive.

Larghezza delle colonne. Consente di modificare l'adattamento automatico della larghezza delle colonne delle tabelle pivot.

- **Adatta per etichette e dati per tabelle di dimensioni piccole o medie.** Nelle tabelle che non superano 10.000 celle, adatta la larghezza della colonna al valore più grande tra l'etichetta della colonna e il valore di dati più grande. Per le tabelle con più di 10.000 celle, adatta la larghezza delle colonne alla larghezza dell'etichetta di colonna.
- **Adatta solo per etichette.** Consente di adattare la larghezza delle colonne a quella delle relative etichette. In questo modo le tabelle sono più compatte ma è possibile che i valori di dati di larghezza superiore a quella dell'etichetta vengano troncati.
- **Adatta per etichette e dati.** Adatta la larghezza della colonna al valore più alto tra l'etichetta della colonna e il valore di dati maggiore. In questo modo si otterranno tabelle più grandi in cui, tuttavia, verranno visualizzati tutti i valori.

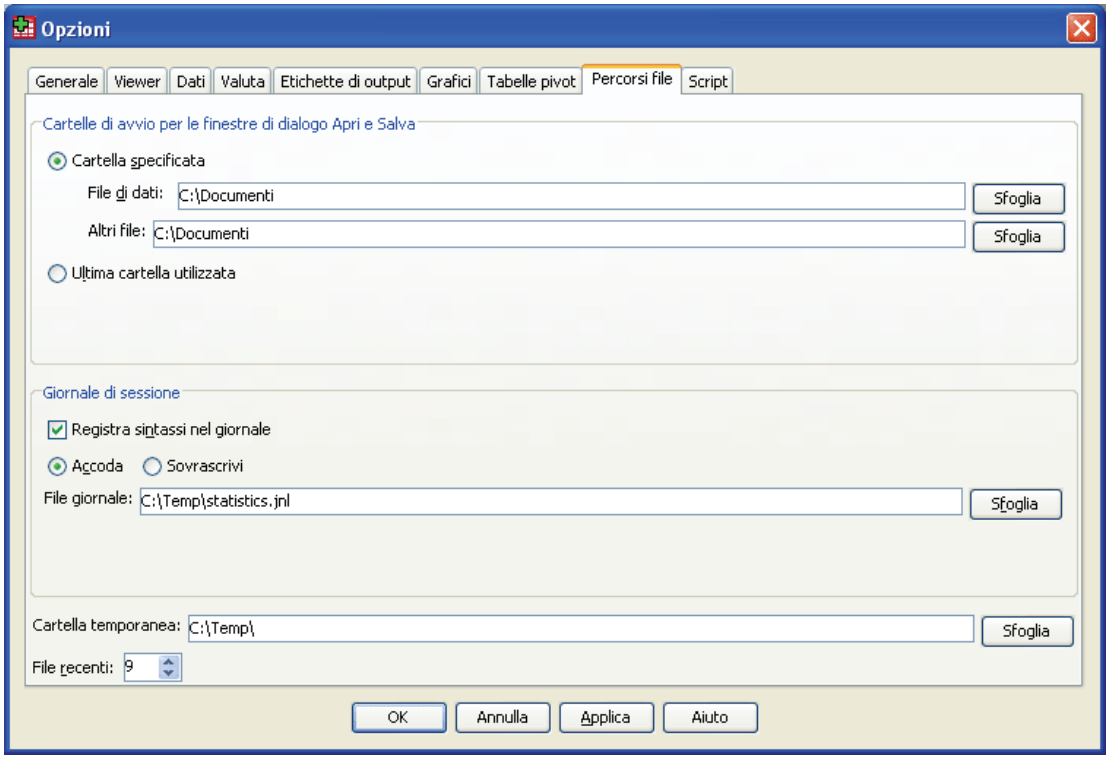
Modalità di modifica predefinita. Consente di impostare l'attivazione delle tabelle pivot nella finestra del Viewer o in una finestra separata. Per impostazione predefinita, facendo doppio clic su una tabella pivot si attiva tutto tranne le tabelle molto grandi presenti nella finestra del Viewer. È possibile attivare le tabelle pivot in una finestra separata oppure selezionare un'impostazione di dimensioni per aprire tabelle pivot più piccole nella finestra del Viewer e tabelle pivot più grandi in una finestra separata.

Copia tabelle grandi negli Appunti in formato RTF. Quando si incollano le tabelle pivot in formato Word/RTF, le tabelle troppo grandi per la larghezza del documento verranno mandate a capo, ridotte di dimensioni per adattarle alla larghezza del documento oppure lasciate invariate.

Opzioni di Percorsi file

Le opzioni della scheda Percorsi file consentono di controllare il percorso predefinito utilizzato dall'applicazione per l'apertura e il salvataggio dei file all'inizio di ogni sessione, il percorso del file giornale, quello della cartella temporanea e il numero dei file visualizzati nell'elenco dei file utilizzati di recente.

Figura 48-9
Finestra di dialogo Opzioni: Scheda Percorsi file



Cartelle di avvio per le finestre di dialogo Apri e Salva

- **Cartella specificata.** La cartella specificata viene utilizzata come percorso predefinito all'inizio di ogni sessione. È possibile specificare diversi percorsi predefiniti per i file di dati e per altri file non di dati.
- **Ultima cartella utilizzata.** L'ultima cartella utilizzata per l'apertura o il salvataggio dei file nella sessione precedente viene impiegata come cartella predefinita all'inizio della sessione successiva. Ciò si applica sia ai file di dati che ad altri file non di dati.

Queste impostazioni si applicano soltanto alle finestre di dialogo per l'apertura e il salvataggio dei file e "l'ultima cartella utilizzata" viene determinata in base all'ultima finestra di dialogo utilizzata per aprire o salvare il file. I file aperti o salvati tramite la sintassi dei comandi non hanno alcun effetto su queste impostazioni e non sono interessati da queste ultime. Queste impostazioni sono disponibili solo in modalità di analisi locale. In modalità di analisi distribuita connessa a un server remoto (richiede SPSS Statistics Server), non è possibile controllare i percorsi delle cartelle di avvio.

Giornale di sessione

È possibile utilizzare il giornale di sessione per registrare automaticamente i comandi eseguiti in una sessione. Vengono registrati anche i comandi inseriti ed eseguiti nelle finestre di sintassi e quelli generati dalle impostazioni delle finestre di dialogo. È possibile modificare il file giornale e

riutilizzare i comandi in altre sessioni di SPSS. È possibile attivare e disattivare la registrazione del giornale, accodare o sovrascrivere il file giornale e selezionare il nome di file e la posizione del giornale. È possibile copiare la sintassi dei comandi dal file giornale e salvarla in un file di sintassi.

Cartella temporanea.

Specifica la posizione dei file temporanei creati nel corso di una sessione. In modalità distribuita, disponibile con la versione server, ciò non influisce sulla posizione dei file di dati temporanei. In modalità distribuita, la posizione dei file di dati temporanei è controllata dalla variabile di ambiente *SPSSTMPDIR*, che può essere impostata esclusivamente nel computer in cui è in esecuzione la versione server del programma. Se si desidera modificare la posizione della directory temporanea, contattare l'amministratore del sistema.

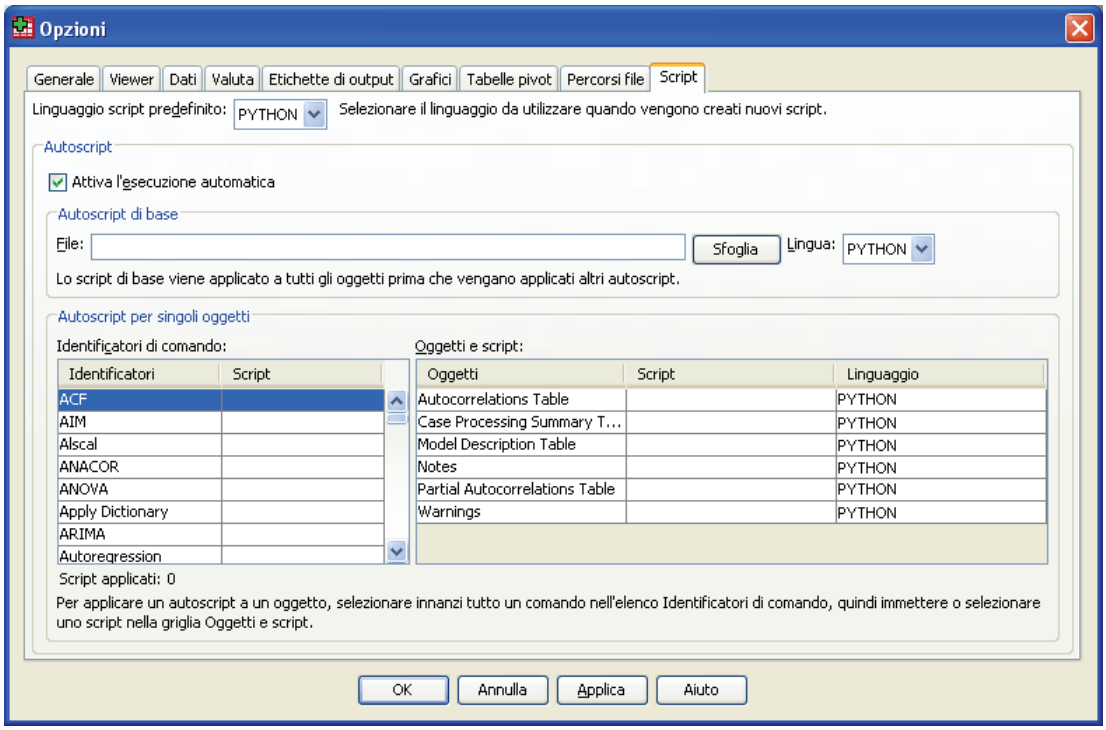
File recenti

Controlla il numero degli ultimi file utilizzati che compaiono nel menu File.

Opzioni: Script

Utilizzare la scheda Script per specificare il linguaggio per script predefinito e qualsiasi script automatico che si desidera impiegare. È possibile utilizzare gli script per eseguire automaticamente numerose funzioni, ad esempio la personalizzazione delle tabelle pivot.

Figura 48-10
Finestra di dialogo Opzioni: Scheda Script



Nota: gli utenti che hanno utilizzato in precedenza Sax Basic devono convertire manualmente qualsiasi script automatico personalizzato. Gli script automatici installati con le versioni precedenti alla 16.0 sono disponibili come insieme di file di script separati che si trova nella sottodirectory *Samples* della directory in cui è installato SPSS Statistics. Per impostazione predefinita, nessun elemento di output è associato agli script automatici. È necessario associare manualmente tutti gli script automatici agli elementi di output desiderati, come illustrato di seguito.

Linguaggio script predefinito. Il linguaggio per script predefinito stabilisce l'editor degli script che deve essere avviato quando vengono creati nuovi script. Specifica inoltre il linguaggio predefinito del file eseguibile che verrà utilizzato per l'esecuzione degli script automatici. I linguaggi per script disponibili dipendono dalla piattaforma in uso. Per Windows, i linguaggi per script disponibili sono Basic, che è installato con il sistema Base e il linguaggio di programmazione Python. Per tutte le altre piattaforme, il linguaggio per script è disponibile con il linguaggio di programmazione Python.

Nota: Per attivare la creazione di script con il linguaggio di programmazione Python, è necessario installare separatamente Python e il plug-in di integrazione SPSS Statistics-Python. Il linguaggio di programmazione Python non viene installato con il modulo Base, ma è disponibile sul CD di installazione per Windows. Il plug-in di integrazione SPSS Statistics-Python è disponibile sul CD di installazione per tutte le piattaforme..

Attiva l'esecuzione automatica. Questa casella di controllo consente di attivare o disattivare l'esecuzione automatica degli script. Per impostazione predefinita, l'esecuzione automatica degli script è attivata.

Autoscript di base. Uno script opzionale che viene applicato a tutti i nuovi oggetti del Viewer prima che venga applicato qualsiasi altro script automatico. Consente di specificare il file di script che deve essere utilizzato come script automatico di base, nonché il linguaggio del file eseguibile che verrà utilizzato per l'esecuzione dello script.

Per applicare gli script automatici agli elementi di output

- Nella griglia Identificatori di comando, selezionare un comando che genera elementi di output ai quali verranno applicati gli script automatici.

  La colonna *Oggetti*, nella griglia Oggetti e script, mostra un elenco degli oggetti associati al comando selezionato. La colonna *Script* mostra qualsiasi script esistente per il comando selezionato.

- Specificare uno script per uno qualsiasi degli elementi mostrati nella colonna *Oggetti*. Fare clic sulla cella *Script* corrispondente. Immettere il percorso allo script oppure fare clic sul pulsante con i puntini (...) per cercare lo script.

- Specificare il linguaggio del file eseguibile che verrà utilizzato per l'esecuzione dello script. *Nota*: se si cambia il linguaggio per script predefinito ciò non avrà effetto sul linguaggio selezionato.

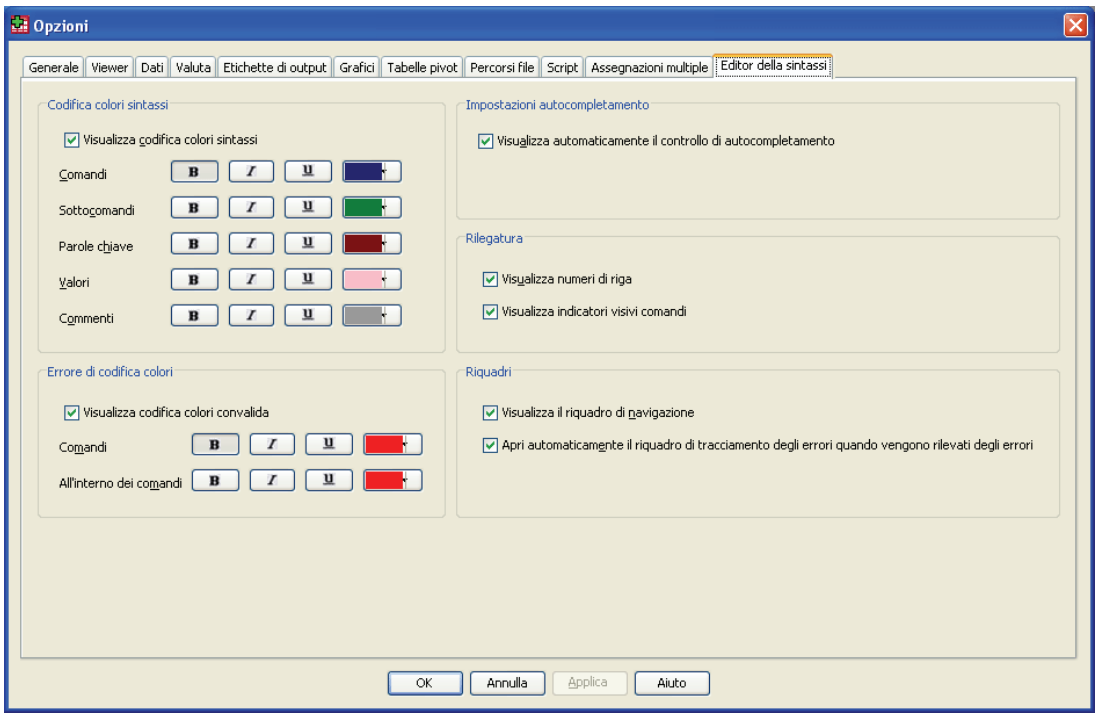
- Fare clic su Applica o su OK.

Per rimuovere le associazioni degli script automatici

- Nella griglia Oggetti e script, fare clic sulla cella della colonna Script corrispondente allo script da dissociare.
- Eliminare il percorso dello script, quindi fare clic su qualsiasi altra cella presente nella griglia Oggetti e script.
- Fare clic su Applica o su OK.

Opzioni Editor della sintassi

Figura 48-11
Finestra di dialogo Opzioni: scheda Editor della sintassi



Codifica colori sintassi

È possibile attivare e disattivare la codifica colori di comandi, sottocomandi, parole chiave, valori di parole chiave e commenti; inoltre, per ognuno di questi elementi, è possibile specificare lo stile del carattere e il colore.

Codifica colori errori

È possibile attivare e disattivare la codifica colori di determinati errori di sintassi e specificare lo stile del carattere e il colore utilizzato. Sia il nome che il testo del comando contenente l'errore vengono codificati in base ai colori ed è possibile scegliere stili diversi per ognuno di questi elementi. Per ulteriori informazioni, vedere Codifica colori in Capitolo 13 a pag. 275.

Impostazioni autocompletamento

La visualizzazione automatica del controllo autocompletamento può essere attivata e disattivata. Per visualizzare in qualsiasi momento il controllo di autocompletamento, premere Ctrl+barra spaziatrice. Per ulteriori informazioni, vedere Autocompletamento in Capitolo 13 a pag. 275.

Rilegatura

Queste opzioni consentono di specificare il comportamento predefinito per visualizzare o nascondere i numeri di riga e gli indicatori visivi dei comandi nella rilegatura dell'Editor della sintassi, vale a dire l'area a sinistra del riquadro di testo riservata a numeri di riga, segnalibri, punti di interruzione e indicatori visivi dei comandi. Gli indicatori visivi dei comandi sono icone che offrono un'indicazione visiva dell'inizio e della fine di un comando.

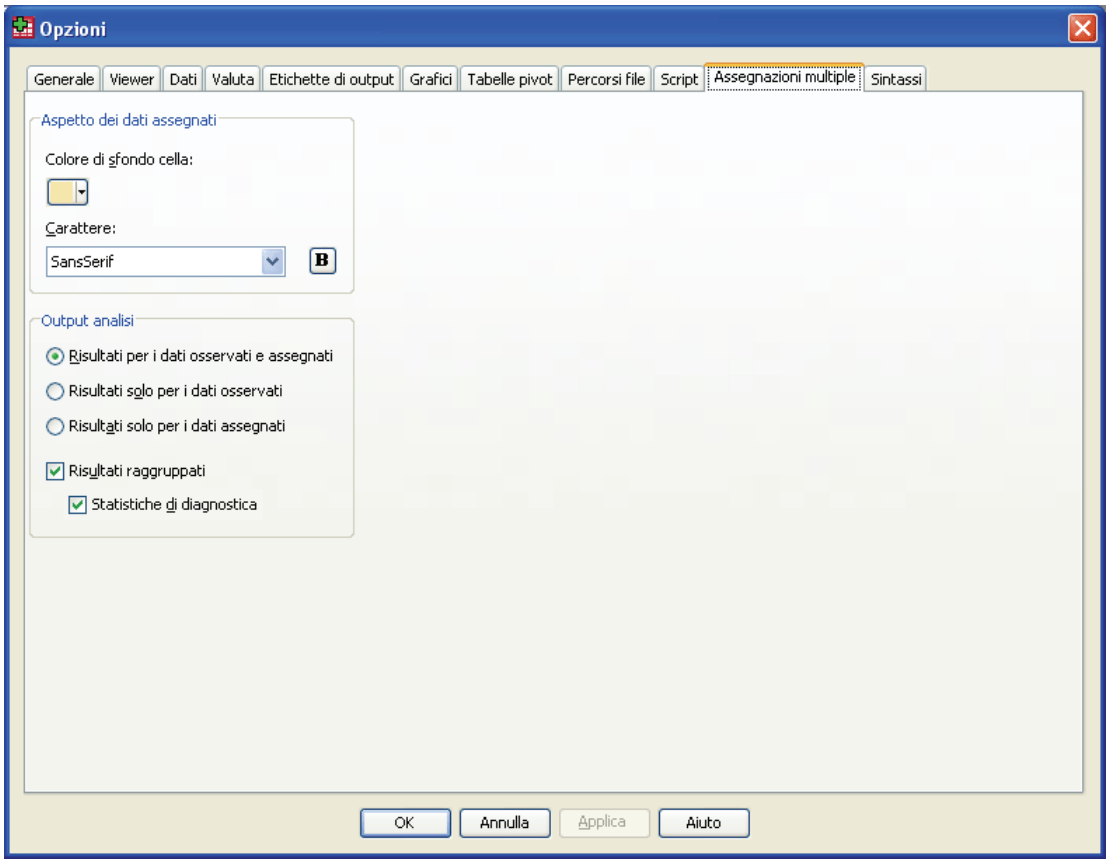
Riquadri

Visualizza il riquadro di navigazione. Questa opzione specifica l'impostazione predefinita per visualizzare o nascondere il riquadro di navigazione. Il riquadro di navigazione contiene un elenco di tutti i comandi riconosciuti nella finestra della sintassi, visualizzati nell'ordine in cui appaiono. Facendo clic su un comando nel riquadro di navigazione, il cursore viene posizionato all'inizio del comando.

Apri automaticamente il riquadro di tracciamento degli errori quando vengono rilevati degli errori. Specifica l'impostazione predefinita per visualizzare o nascondere il riquadro di tracciamento degli errori quando vengono rilevati degli errori di run-time.

Opzioni Assegnazioni multiple

Figura 48-12
Finestra di dialogo Opzioni: Scheda Assegnazioni multiple



La scheda Assegnazioni multiple consente di controllare due tipi di preferenze relative alle assegnazioni multiple:

Aspetto dei dati assegnati. Per impostazione predefinita, le celle che contengono i dati assegnati hanno uno sfondo di colore diverso rispetto alle celle contenenti dati non assegnati. L'aspetto facilmente distinguibile dei dati assegnati dovrebbe facilitare la consultazione di un insieme di dati e l'individuazione di tali celle. È possibile cambiare il colore predefinito dello sfondo della cella e il carattere, nonché visualizzare i dati assegnati in grassetto.

Output analisi. Questo gruppo controlla il tipo di output del Viewer prodotto ogni volta che viene analizzato un insieme di dati ad assegnazione multipla. Per impostazione predefinita, verrà prodotto output per l'insieme di dati originale (precedente all'assegnazione) e per ogni insieme di dati assegnato. Inoltre, nel caso delle procedure che supportano il raggruppamento in pool dei dati assegnati, verranno generati risultati raggruppati finali. Quando viene eseguito il raggruppamento univariato, viene visualizzata anche la diagnostica del raggruppamento. Tuttavia, è possibile sopprimere qualsiasi output che non si desidera produrre.

Per specificare le impostazioni di assegnazione multipla

Dai menu, scegliere:

Modifica
Opzioni

Fare clic sulla scheda Assegnazione multipla.

Capitolo
49

Personalizzazione dei menu e delle barre degli strumenti

Editor dei menu

È possibile utilizzare l'Editor dei menu per personalizzare i menu. Nell'Editor dei menu è possibile:

- Aggiungere comandi di menu per l'esecuzione di script personalizzati.
- Aggiungere comandi di menu per l'esecuzione di file della sintassi dei comandi.
- Aggiungere comandi di menu per l'avvio di altre applicazioni e l'invio automatico di dati ad altre applicazioni.

È possibile inviare dati ad altre applicazioni nei seguenti formati: SPSS Statistics, Excel, Lotus 1-2-3, delimitato da tabulazioni e dBASE IV.

Per aggiungere comandi ai menu

▶ Dai menu, scegliere:
Visualizza
Editor del menu...

▶ Nella finestra di dialogo Editor dei menu, fare doppio clic sul menu (o fare clic sul segno più) a cui si desidera aggiungere un nuovo comando.

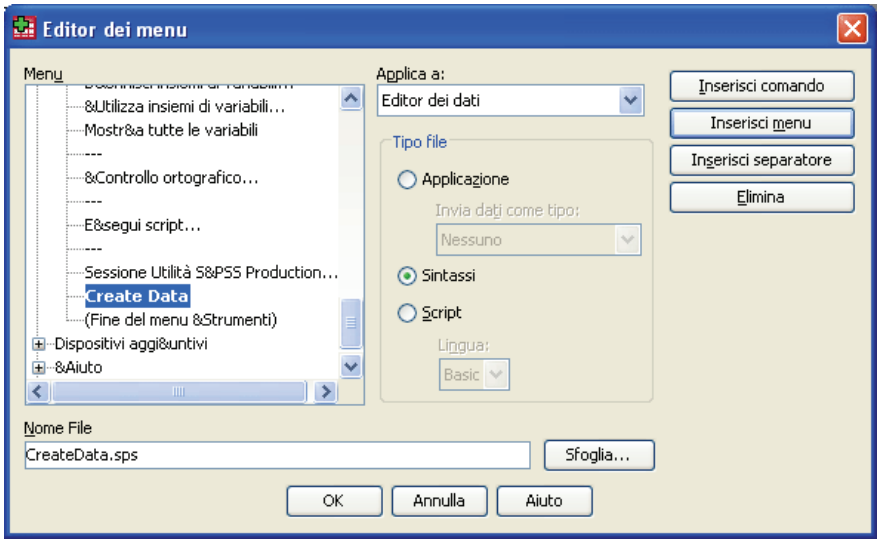
▶ Selezionare il comando del menu che si desidera preceda il nuovo comando.

▶ Fare clic su Inserisci comando per inserire un nuovo comando nel menu.

▶ Selezionare il tipo di file per il nuovo comando (file script, file della sintassi dei comandi o applicazione esterna).

▶ Fare clic su Sfoglia per selezionare un file da associare al nuovo comando di menu.

Figura 49-1
Finestra di dialogo Editor dei menu



È inoltre possibile aggiungere menu completamente nuovi e separatori tra i comandi dei menu, nonché inviare automaticamente il contenuto dell'Editor dei dati a un'altra applicazione selezionandola dal menu.

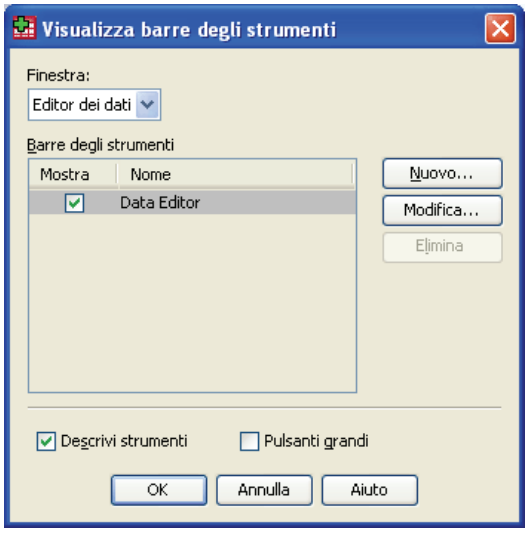
Personalizzazione delle barre degli strumenti

È possibile personalizzare le barre degli strumenti e crearne di nuove. Le barre degli strumenti possono contenere tutti gli strumenti disponibili, inclusi quelli per le azioni dei menu. Possono inoltre esservi inclusi strumenti personalizzati che permettono di avviare altre applicazioni, eseguire file di sintassi o eseguire file di script.

Visualizza barre degli strumenti

La finestra Visualizza barre degli strumenti consente di visualizzare o nascondere le barre degli strumenti, personalizzarle oppure crearne di nuove. Le barre degli strumenti possono contenere tutti gli strumenti disponibili, inclusi quelli per le azioni dei menu. Possono inoltre esservi inclusi strumenti personalizzati che permettono di avviare altre applicazioni, eseguire file di sintassi o eseguire file di script.

Figura 49-2
Finestra di dialogo Visualizza barre degli strumenti



Per personalizzare le barre degli strumenti

- Dai menu, scegliere:

  Visualizza
  Barre degli strumenti
  Personalizza

- Selezionare la barra degli strumenti che si desidera personalizzare e quindi fare clic su Modifica oppure su Nuova barra per creare una nuova barra degli strumenti.
- Per le nuove barre degli strumenti, specificare il nome, selezionare le finestre in cui si desidera inserire la nuova barra e quindi fare clic su Modifica.
- Selezionare una voce dell'elenco Categorie per visualizzare gli strumenti disponibili nella categoria desiderata.
- Trascinare gli strumenti desiderati sulla barra degli strumenti visualizzata nella finestra di dialogo.
- Per rimuovere uno strumento, trascinarlo all'esterno della barra degli strumenti visualizzata nella finestra di dialogo.

  Per creare uno strumento personalizzato per l'apertura di file, l'esecuzione di file di sintassi o di script:

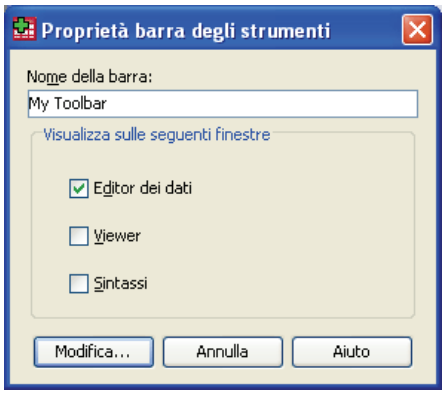
- Fare clic su Nuovo strumento nella finestra di dialogo Modifica barra degli strumenti.
- Digitare un'etichetta descrittiva per lo strumento.
- Selezionare il comando che si desidera assegnare allo strumento (apertura di file, esecuzione di file di sintassi o di script).
- Fare clic su Sfoglia per selezionare un file o un'applicazione da associare allo strumento.

I nuovi strumenti verranno visualizzati nella categoria Personalizzati, che contiene anche i comandi di menu definiti dall'utente.

Proprietà barra degli strumenti

Nella finestra di dialogo Proprietà barra degli strumenti è possibile selezionare i tipi di finestra in cui si desidera visualizzare la barra degli strumenti selezionata. In questa finestra di dialogo è inoltre possibile creare i nomi delle nuove barre degli strumenti.

Figura 49-3
Finestra di dialogo Proprietà barra degli strumenti



Per impostare le proprietà delle barre degli strumenti

- Dai menu, scegliere:
  Visualizza
  Barre degli strumenti
  Personalizza

- Per le barre degli strumenti esistenti, fare clic su Modifica e quindi su Proprietà nella finestra di dialogo Modifica barra degli strumenti.

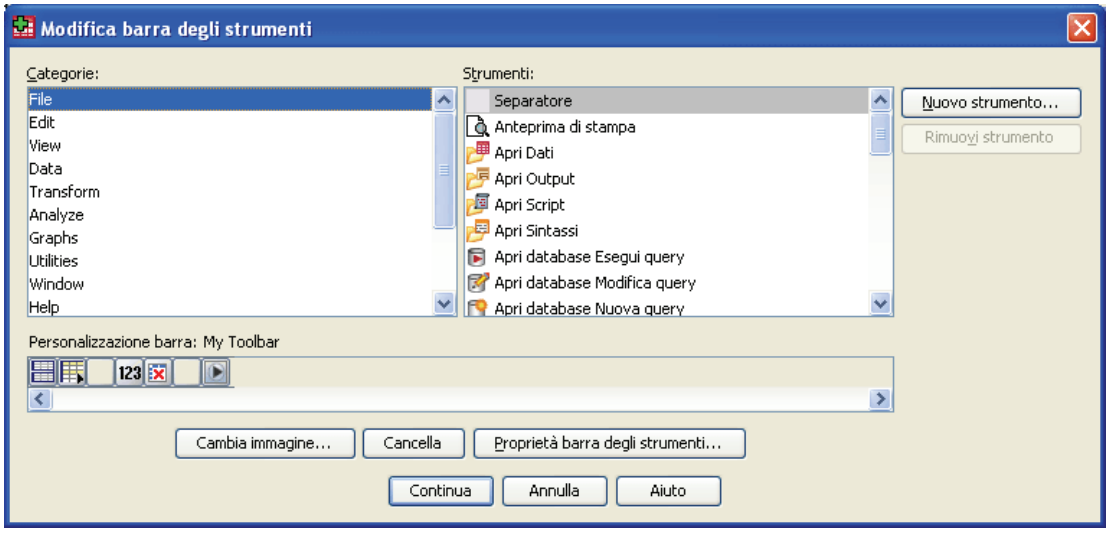
- Per le nuove barre degli strumenti, fare clic su Nuovo strumento.

- Selezionare i tipi di finestra in cui si desidera visualizzare la barra degli strumenti. Per le nuove barre degli strumenti, specificare anche il nome.

Modifica barra degli strumenti

Nella finestra di dialogo Modifica barra degli strumenti è possibile creare nuove barre degli strumenti e personalizzare quelle esistenti. Le barre degli strumenti possono contenere tutti gli strumenti disponibili, inclusi quelli per le azioni dei menu. Possono inoltre esservi inclusi strumenti personalizzati che permettono di avviare altre applicazioni, eseguire file di sintassi o eseguire file di script.

Figura 49-4
Finestra di dialogo Personalizza barra degli strumenti



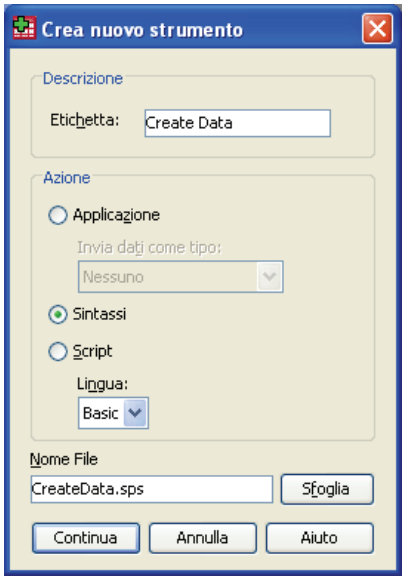
Per cambiare le immagini delle barre degli strumenti

- Selezionare lo strumento della barra degli strumenti di cui si desidera cambiare l'immagine.
- Fare clic su Cambia immagine.
- Selezionare il file di immagine da utilizzare per lo strumento. I formati immagine disponibili supportati sono: BMP, PNG, GIF e JPG.
  - Le immagini devono essere di forma quadrata. Altre forme di immagini vengono ritagliate in un'area quadrate.
  - Le immagini vengono adattate automaticamente. Per la visualizzazione ottimale, utilizzare il formato 16x16 pixel per le immagini piccole della barra degli strumenti oppure il formato 32x32 pixel per le immagini grandi.

Crea nuovo strumento

Nella finestra di dialogo Crea nuovo strumento è possibile creare strumenti personalizzati per l'avvio di altre applicazioni, eseguire file di sintassi e file di script.

Figura 49-5
Finestra di dialogo Crea nuovo strumento



Capitolo
50

Creazione e gestione di finestre di dialogo personalizzate

Il Generatore di finestre di dialogo personalizzate consente di creare e gestire le finestre di dialogo personalizzate per la generazione della sintassi dei comandi. Utilizzando il Generatore di finestre di dialogo personalizzate è possibile:

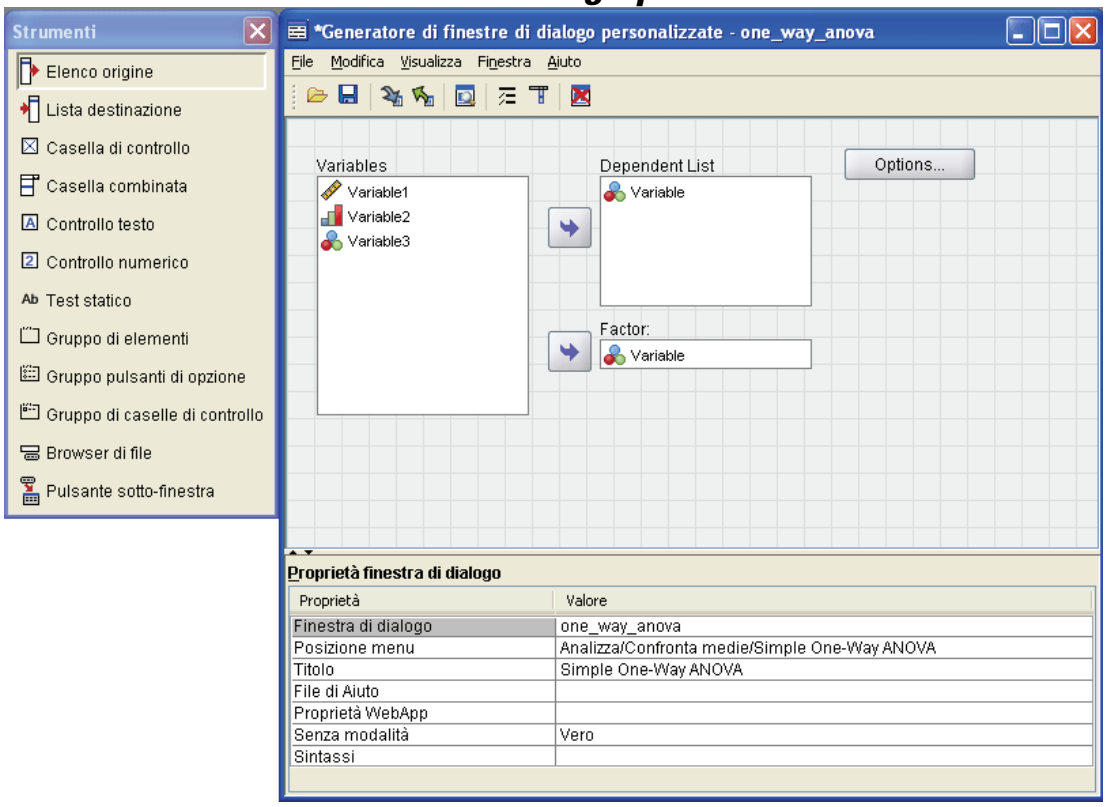
- Creare una versione personalizzata di una finestra di dialogo per una procedura SPSS Statistics incorporata. Ad esempio, è possibile creare una finestra di dialogo per la procedura Frequenze che consente all'utente di selezionare l'insieme di variabili e quindi genera la sintassi dei comandi con opzioni preimpostate che standardizzano l'output.
- Creare un'interfaccia utente che genera la sintassi di un comando di estensione. I comandi di estensione sono comandi SPSS Statistics definiti dall'utente che vengono implementati nel linguaggio di programmazione Python o R. Per ulteriori informazioni, vedere Finestre di dialogo personalizzate per i comandi di estensione a pag. 572.
- Aprire un file contenente la specifica di una finestra di dialogo personalizzata, magari creata da un altro utente, quindi aggiungerla alla propria installazione di SPSS Statistics, apportando le proprie modifiche, se si desidera.
- Salvare la specifica in modo che altri utenti possano aggiungerla alle loro installazioni di SPSS Statistics.

Come avviare il Generatore di finestre di dialogo personalizzate

▶ Dai menu, scegliere:

Strumenti
  Generatore di finestre di dialogo personalizzate

Layout del Generatore di finestre di dialogo personalizzate



Disegno

L'area di disegno è la zona del Generatore di finestre di dialogo personalizzate in cui si progetta il layout della finestra di dialogo.

Riquadro delle proprietà

Il riquadro delle proprietà è l'area del Generatore di finestre di dialogo personalizzate in cui si specificano le proprietà dei controlli della finestra di dialogo, nonché della finestra di dialogo stessa, quali la posizione dei menu.

Tavolozza strumenti

La tavolozza strumenti fornisce la serie di controlli che possono essere inclusi in una finestra di dialogo personalizzata. È possibile visualizzare o nascondere la tavolozza strumenti selezionando Tavolozza strumenti dal menu Visualizza.

Creazione di una finestra di dialogo personalizzata

I passaggi principali per la creazione di una finestra di dialogo personalizzata sono:

- Specificare le proprietà della finestra di dialogo stessa, quale il titolo visualizzato all'apertura della finestra e la posizione del nuovo comando di menu per la finestra all'interno dei menu di SPSS Statistics. Per ulteriori informazioni, vedere Proprietà della finestra di dialogo a pag. 551.
- Specificare i controlli, quali gli elenchi delle variabili sorgente e di destinazione, che compongono la finestra di dialogo e le sotto-finestre. Per ulteriori informazioni, vedere Tipi di controlli a pag. 560.
- Creare il modello di sintassi che specifica la sintassi dei comandi che la finestra di dialogo deve generare. Per ulteriori informazioni, vedere Creazione del modello di sintassi a pag. 555.
- Installare la finestra di dialogo su SPSS Statistics e/o salvare la specifica della finestra di dialogo in un file pacchetto finestra di dialogo personalizzato (*.spd*). Per ulteriori informazioni, vedere Gestione delle finestre di dialogo personalizzate a pag. 558.

  Durante la creazione della finestra di dialogo è possibile visualizzarne un'anteprima. Per ulteriori informazioni, vedere Visualizzazione dell'anteprima di una finestra di dialogo personalizzata a pag. 557.

Proprietà della finestra di dialogo

Per visualizzare e impostare le Proprietà finestra di dialogo:

- Fare clic sull'area di disegno al di fuori dei controlli. Senza alcun controllo sull'area di disegno, le Proprietà finestra di dialogo sono sempre visibili.

  Finestra di dialogo. La proprietà Finestra di dialogo è obbligatoria e consente di specificare un nome univoco da associare alla finestra. Questo è il nome utilizzato per identificare la finestra di dialogo durante l'installazione o la disinstallazione. Per ridurre la possibilità di conflitti tra i nomi, conviene anteporre un prefisso al nome della finestra con un identificatore della propria azienda, ad esempio un URL.

  Posizione menu. Fare clic sul pulsante con i puntini (...) per aprire la finestra di dialogo Posizione menu, che consente di specificare il nome e la posizione del comando di menu per la finestra di dialogo. Per aggiungere una posizione di menu per una nuova finestra di dialogo personalizzata o modificare la posizione di menu di una finestra di dialogo personalizzata installata è necessario riavviare SPSS Statistics.

  Titolo. La proprietà Titolo specifica il testo da visualizzare nella barra del titolo della finestra di dialogo.

  File di Aiuto. La proprietà File di Aiuto è facoltativa e specifica il percorso a un file di Aiuto relativo alla finestra di dialogo. È il file che verrà avviato quando l'utente fa clic sul pulsante Aiuto nella finestra di dialogo. I file di Aiuto devono essere in formato HTML. Il pulsante Aiuto sulla finestra di dialogo in fase di esecuzione è nascosto se non è associato a un file di Aiuto.

Tutti i file di supporto, quali i file di immagine e i fogli di stile, devono essere memorizzati con il file di Aiuto principale quando la finestra di dialogo è stata installata. Per impostazione predefinita, le specifiche di una finestra di dialogo personalizzata installata vengono memorizzate nella cartella *ext/lib/<Finestra di dialogo>* della directory di installazione, che si trova nella directory principale della directory di installazione per Windows e Linux e nella cartella *Contents/bin* nel bundle dell'applicazione per Mac. I file di supporto devono trovarsi nella directory principale della cartella e non all'interno di sottocartelle. Devono inoltre essere aggiunti manualmente ai file pacchetto personalizzati creati per la finestra di dialogo.

Se sono state specificate posizioni alternative per le finestre di dialogo installate (utilizzando la variabile di ambiente *SPSS_CDIALOGS_PATH*), memorizzare tutti i file di supporto nella cartella *<Finestra di dialogo>* nella posizione alternativa corrispondente. Per ulteriori informazioni, vedere Gestione delle finestre di dialogo personalizzate a pag. 558.

Nota: quando si lavora su una finestra di dialogo aperta da un file pacchetto finestra di dialogo personalizzato (*.spd*), la proprietà File di Aiuto punta a una cartella temporanea associata al file *.spd*. Tutte le modifiche al file di Aiuto devono essere apportate alla copia nella cartella temporanea.

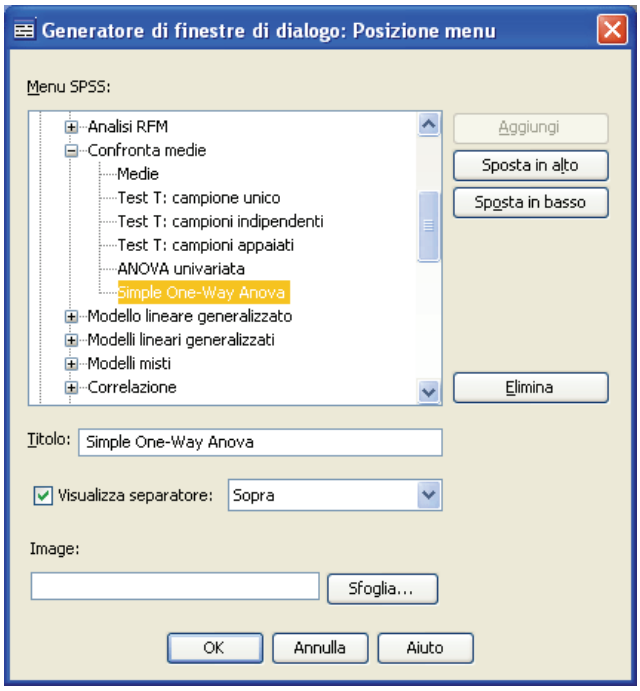
Proprietà WebApp. Consente di associare un file delle proprietà a questa finestra di dialogo, da utilizzare per creare applicazioni thin client sviluppate sul Web.

Senza modalità. Specifica se la finestra di dialogo è modale o senza modalità. Quando una finestra di dialogo è modale, deve essere chiusa prima che l'utente possa interagire con le finestre dell'applicazione principale (Dati, Output e Sintassi) o con altre finestre di dialogo aperte. Le finestre di dialogo senza modalità non sono soggette a tale vincolo. L'impostazione predefinita è senza modalità.

Sintassi. La proprietà Sintassi specifica il modello della sintassi, utilizzato per creare la sintassi dei comandi generata dalla finestra di dialogo in fase di esecuzione. Fare clic sul pulsante con i puntini (...) per aprire il Modello di sintassi. Per ulteriori informazioni, vedere Creazione del modello di sintassi a pag. 555.

Specifica della posizione dei menu per una finestra di dialogo personalizzata

Figura 50-1
Finestra di dialogo Posizione menu del Generatore di finestre di dialogo personalizzate



La finestra di dialogo Posizione menu consente di specificare il nome e la posizione del comando di menu di una finestra di dialogo. Per aggiungere una posizione di menu per una nuova finestra di dialogo personalizzata o modificare la posizione di menu di una finestra di dialogo personalizzata installata è necessario riavviare SPSS Statistics. I comandi di menu delle finestre di dialogo personalizzate non appaiono nell'Editor dei menu di SPSS Statistics.

- Fare doppio clic sul menu (o fare clic sul segno più) a cui si desidera aggiungere il comando per la nuova finestra di dialogo. È anche possibile aggiungere comandi al menu di livello superiore denominato Personalizzato (posizionato tra i comandi Grafici e Strumenti), disponibile solo per i comandi di menu associati alle finestre di dialogo personalizzate.

  Se si desidera creare menu o sottomenu personalizzati, utilizzare l'Editor dei menu. Per ulteriori informazioni, vedere Editor dei menu in Capitolo 49 a pag. 543. Si noti tuttavia che gli altri utenti della finestra di dialogo dovranno creare manualmente lo stesso menu o sottomenu dall'Editor dei menu, altrimenti la finestra verrà aggiunta al loro menu Personalizzato.

  Nota: la finestra di dialogo Posizione menu visualizza tutti i menu, inclusi quelli di tutti i moduli aggiuntivi. Assicurarsi di aggiungere la voce di menu per la finestra di dialogo personalizzata a un menu che sarà disponibile all'utente o agli utenti della finestra di dialogo.

- Selezionare il comando di menu che si desidera preceda la nuova finestra di dialogo. Una volta aggiunto il comando è possibile utilizzare i pulsanti Sposta in alto e Sposta in basso per riposizionarlo.

- Immettere un titolo per il comando di menu. I titoli all'interno di un dato menu o sottomenu devono essere univoci.
- Fare clic su Aggiungi.

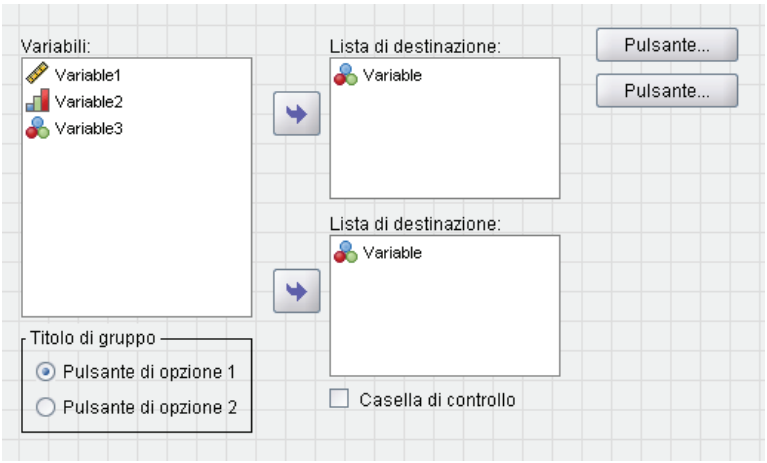
Se lo si desidera, è possibile:

- Aggiungere un separatore sopra o sotto il nuovo comando di menu.
- Specificare il percorso a un'immagine che verrà visualizzata accanto al comando di menu per la finestra di dialogo. I tipi di immagine supportati sono *gif* e *png*. L'immagine non può avere dimensioni superiori a 16 x 16 pixel.

Layout dei controlli sull'area di disegno

Per aggiungere controlli a una finestra di dialogo personalizzata, trascinarli dalla tavolozza strumenti all'area di disegno. Per garantire l'uniformità con le finestre di dialogo incorporate, l'area di disegno è divisa in tre colonne funzionali nelle quali è possibile posizionare i controlli.

Figura 50-2
Struttura dell'area di disegno



Colonna sinistra. La colonna sinistra è destinata principalmente a un controllo elenco origine. Nella colonna sinistra possono essere posizionati tutti i controlli diversi dalle liste destinazione e dai pulsanti delle sotto-finestre.

Colonna centrale. La colonna centrale può contenere qualsiasi controllo diverso dagli elenchi origine e dai pulsanti delle sotto-finestre.

Colonna destra. La colonna destra può contenere solo pulsanti delle sotto-finestre.

Anche se non sono mostrati nell'area di disegno, la parte inferiore di ogni finestra di dialogo personalizzata contiene pulsanti OK, Incolla, Annulla e Aiuto. La presenza e la posizione di tali pulsanti è automatica, tuttavia il pulsante Aiuto è nascosto se non esiste un file di Aiuto associato alla finestra di dialogo (specificato dalla proprietà File di Aiuto in Proprietà finestra di dialogo).

È possibile modificare l'ordine verticale dei controlli all'interno di una colonna trascinandoli su o giù, ma l'esatta posizione dei controlli verrà determinata automaticamente. Al run-time, i controlli verranno ridimensionati nel modo appropriato quando viene ridimensionata la finestra di dialogo

stessa. Controlli quali gli elenchi origine e le liste destinazione si espandono automaticamente per riempire lo spazio sottostante disponibile.

Creazione del modello di sintassi

Il modello di sintassi specifica la sintassi dei comandi che verrà generata dalla finestra di dialogo. Un'unica finestra di dialogo personalizzata può generare la sintassi dei comandi per qualsiasi numero di comandi SPSS Statistics incorporati o di estensione.

Il modello di sintassi può essere composto da testo statico che viene sempre generato e da identificatori di controllo sostituiti in fase di esecuzione con i valori dei controlli associati della finestra di dialogo personalizzata. Ad esempio, i nomi dei comandi e le specifiche dei sottocomandi che non dipendono dall'input dell'utente saranno testo statico, mentre l'insieme di variabili specificate in una lista destinazione sarà rappresentato dall'identificatore di controllo per il controllo lista destinazione.

Per creare il modello di sintassi

- Dai menu del Generatore di finestre di dialogo personalizzate selezionare:

  Modifica
  Modello di sintassi

  (Oppure fare clic sul pulsante con i puntini (...) nel campo della proprietà Sintassi in Proprietà finestra di dialogo)

- Per la sintassi dei comandi statici che non dipende da valori specificati dall'utente, immettere la sintassi allo stesso modo in cui viene immessa nell'Editor della sintassi.

- Aggiungere gli identificatori di controllo di tipo `%%Identificatore%%` nelle posizioni in cui si desidera inserire la sintassi dei comandi generata dai controlli, dove `Identificatore` è il valore della proprietà Identificatore per il controllo. È possibile scegliere in un elenco di identificatori di controllo disponibili premendo Ctrl+barra spaziatrice. Se si immettono gli identificatori manualmente, mantenere gli eventuali spazi, poiché tutti gli spazi sono significativi negli identificatori.

  Per quanto riguarda tutti i controlli diversi dalle caselle di controllo e dai gruppi di caselle di controllo, ogni identificatore viene sostituito dal valore corrente della proprietà Sintassi del controllo associato in fase di esecuzione. Per quanto riguarda le caselle di controllo e i gruppi di caselle di controllo, l'identificatore viene sostituito dal valore corrente della proprietà Sintassi verificata e Sintassi non verificata del controllo associato, a seconda dello stato corrente del controllo (selezionato o deselezionato). Per ulteriori informazioni, vedere Tipi di controlli a pag. 560.

Esempio: inclusione di valori di runtime nel modello di sintassi

Si consideri una versione semplificata della finestra di dialogo Frequenze, che contiene solo un controllo elenco origine e un controllo lista destinazione e genera la sintassi dei comandi nella forma seguente:

```
FREQUENCIES VARIABLES=var1 var2...
```

```
  /FORMAT = NOTABLE
  /BARCHART.
```

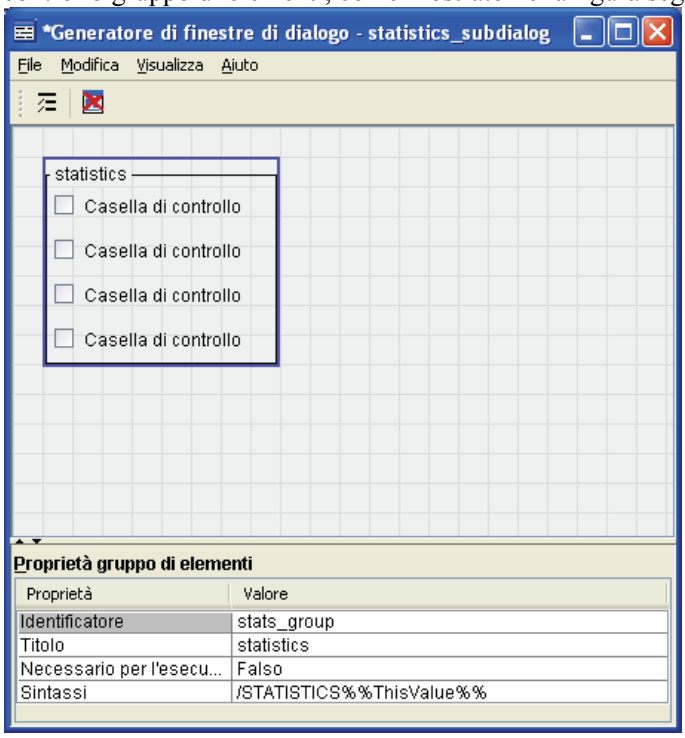
Il modello di sintassi relativo potrebbe avere l'aspetto seguente:

```
FREQUENCIES VARIABLES=%%elenco_destinazione%%
  /FORMAT = NOTABLE
  /BARCHART.
```

- `%%elenco_destinazione%%` è il valore della proprietà Identificatore per il controllo lista destinazione. In fase di esecuzione verrà sostituito dal valore corrente della proprietà Sintassi del controllo.
- Definendo la proprietà Sintassi del controllo lista destinazione in modo che sia `%%QuestoValore%%` si specifica che, in fase di esecuzione, il valore corrente della proprietà sarà il valore del controllo, ovvero l'insieme di variabili nella lista destinazione.

Esempio: inclusione di sintassi dei comandi contenitore

Ampliando l'esempio seguente, si consideri l'aggiunta di una sotto-finestra Statistiche contenente un unico gruppo di caselle di controllo che consentono a un utente di specificare media, deviazione standard, minimo e massimo. Si presuma che le caselle di controllo siano contenute in un controllo gruppo di elementi, come mostrato nella figura seguente.



Segue un esempio della sintassi dei comandi generata:

```
FREQUENCIES VARIABLES=var1 var2...
  /FORMAT = NOTABLE
  /STATISTICS MEAN STDDEV
  /BARCHART.
```

Il modello di sintassi relativo potrebbe avere l'aspetto seguente:

```
FREQUENCIES VARIABLES=%%elenco_destinazione%%
  /FORMAT = NOTABLE
  %%gruppo_stat%%
  /BARCHART.
```

- `%%elenco_destinazione%%` è il valore della proprietà Identificatore per il controllo lista destinazione e `%%gruppo_stat%%` è il valore della proprietà Identificatore per il controllo gruppo di elementi.

La tabella seguente mostra un metodo per specificare le proprietà Sintassi del gruppo di elementi e delle caselle di controllo contenute al suo interno per generare il risultato desiderato. La proprietà Sintassi della lista destinazione deve essere impostata su `%%QuestoValore%%`, come descritto nell'esempio precedente.

Proprietà Sintassi del gruppo di elementi	`/STATISTICS %%QuestoValore%%`
Proprietà Sintassi verificata della casella di controllo media	`MEAN`
Proprietà Sintassi verificata della casella di controllo devst	`STDDEV`
Proprietà Sintassi verificata della casella di controllo min	`MINIMUM`
Proprietà Sintassi verificata della casella di controllo max	`MAXIMUM`

In fase di esecuzione `%%gruppo_stat%%` verrà sostituito dal valore corrente della proprietà Sintassi del controllo gruppo di elementi. Più precisamente, `%%QuestoValore%%` verrà sostituito da un elenco di valori delimitati da spazi vuoti della proprietà Sintassi verificata o Sintassi non verificata di ogni casella di controllo, a seconda dello stato relativo (selezionata o deselezionata). Poiché sono stati specificati valori solo per la proprietà Sintassi verificata, solo le caselle selezionate contribuiranno a `%%QuestoValore%%`. Ad esempio, se l'utente fa clic sulle caselle media e deviazione standard, il valore di runtime della proprietà Sintassi per il gruppo di elementi sarà `/STATISTICS MEAN STDDEV`.

Se non viene selezionata alcuna casella, la proprietà Sintassi per il controllo gruppo di elementi sarà vuota e la sintassi dei comandi generata non conterrà alcun riferimento a `%%gruppo_stat%%`. Questo può essere considerato positivo o negativo. Ad esempio, anche quando non è selezionata alcuna casella si può desiderare di generare ugualmente il comando `STATISTICS`. Ciò può essere ottenuto facendo riferimento agli identificatori per le caselle di controllo direttamente nel modello di sintassi, ad esempio:

```
FREQUENCIES VARIABLES=%%elenco_destinazione%%
  /FORMAT = NOTABLE
  /STATISTICS %%media_stat%% %%devst_stat%% %%min_stat%% %%max_stat%%
  /BARCHART.
```

Visualizzazione dell'anteprima di una finestra di dialogo personalizzata

È possibile visualizzare un'anteprima della finestra di dialogo attualmente aperta nel Generatore di finestre di dialogo personalizzate. La finestra di dialogo viene visualizzata e funziona come se fosse eseguita dai menu di SPSS Statistics.

- Gli elenchi di variabili sorgente sono popolati da variabili fittizie che possono essere spostate nelle liste destinazione.
- Il pulsante Incolla incolla la sintassi dei comandi nella finestra della sintassi designata.
- Il pulsante OK chiude l'anteprima.
- Se viene specificato un file di Aiuto, il pulsante Aiuto è abilitato e aprirà il file specificato. Se non è selezionato alcun file di Aiuto, il pulsante Aiuto sarà disabilitato durante l'anteprima e sarà nascosto quando la finestra di dialogo effettiva verrà eseguita.

Per visualizzare un'anteprima di una finestra di dialogo:

▶ Dai menu del Generatore di finestre di dialogo personalizzate selezionare:
File
Anteprima finestra di dialogo

Gestione delle finestre di dialogo personalizzate

Il Generatore di finestre di dialogo personalizzate consente di gestire le proprie finestre di dialogo o quelle create da altri utenti. È possibile installare, disinstallare o modificare le finestre di dialogo installate, nonché salvare le specifiche di una finestra di dialogo in un file esterno oppure aprire un file contenente le specifiche di una finestra di dialogo per modificarlo. Le finestre di dialogo personalizzate devono essere installate prima di poter essere utilizzate.

Installazione di una finestra di dialogo personalizzata

È possibile installare la finestra di dialogo attualmente aperta nel Generatore di finestre di dialogo personalizzate oppure installarla da un file pacchetto finestra di dialogo personalizzato (*.spd*). La reinstallazione di una finestra di dialogo esistente sostituirà la versione esistente.

Per installare la finestra di dialogo attualmente aperta:

▶ Dai menu del Generatore di finestre di dialogo personalizzate selezionare:
File
Installa

Per installare da un file pacchetto finestra di dialogo personalizzato:

▶ Dai menu, scegliere:
Strumenti
Installa finestra di dialogo

Per utilizzare una nuova finestra di dialogo, sarà necessario riavviare SPSS Statistics. La reinstallazione di una finestra di dialogo esistente non necessita di riavvio, a meno che non venga modificata la posizione del menu.

Per impostazione predefinita, l'installazione di una finestra di dialogo richiede l'autorizzazione di scrittura sulla directory di installazione di SPSS Statistics. Se non si dispone di autorizzazioni di scrittura su tale directory o se si desidera memorizzare le finestre di dialogo installate in un'altra posizione, è possibile specificare una o più posizioni alternative definendo la variabile di ambiente *SPSS_CDIALOGS_PATH*. Se presenti, i percorsi specificati in *SPSS_CDIALOGS_PATH* hanno la

precedenza sulla directory di installazione di SPSS Statistics. Per posizioni multiple, separarle con il punto e virgola su Windows e con i due punti su Linux e Mac.

Apertura di una finestra di dialogo installata

È possibile aprire una finestra di dialogo personalizzata già installata per modificarla e/o salvarla esternamente in modo da poterla distribuire ad altri utenti.

▶ Dai menu del Generatore di finestre di dialogo personalizzate selezionare:
File
Apri installata

Nota: se si apre una finestra di dialogo installata per modificarla, selezionando File > Installa la si reinstalla, sostituendo la versione esistente.

Disinstallazione di una finestra di dialogo personalizzata

▶ Dai menu del Generatore di finestre di dialogo personalizzate selezionare:
File
Disinstalla

Salvataggio in un file pacchetto finestra di dialogo personalizzato

È possibile salvare le specifiche di una finestra di dialogo in un file esterno per distribuire la finestra ad altri utenti o salvare le specifiche di una finestra di dialogo non ancora installata. Le specifiche vengono salvate in un file pacchetto finestra di dialogo personalizzato (*.spd*).

▶ Dai menu del Generatore di finestre di dialogo personalizzate selezionare:
File
Salva

Apertura di un file pacchetto finestra di dialogo personalizzato

È possibile aprire un file pacchetto finestra di dialogo personalizzato contenente le specifiche di una finestra personalizzata per modificarla, salvarla nuovamente o installarla.

▶ Dai menu del Generatore di finestre di dialogo personalizzate selezionare:
File
Apri

Copia manuale di una finestra di dialogo installata

Per impostazione predefinita, le specifiche di una finestra di dialogo personalizzata installata vengono memorizzate nella cartella *ext/lib/<Finestra di dialogo>* della directory di installazione, che si trova nella directory principale della directory di installazione per Windows e Linux e nella cartella *Contents/bin* nel bundle dell'applicazione per Mac. È possibile copiare questa cartella nella stessa posizione relativa in un'altra istanza di SPSS Statistics: verrà riconosciuta come finestra di dialogo installata al successivo avvio di tale istanza.

- Se sono state specificate posizioni alternative per le finestre di dialogo installate (utilizzando la variabile di ambiente *SPSS_CDIALOGS_PATH*), copiare la cartella <*Finestra di dialogo*> dalla posizione alternativa corrispondente.
- Se sono state definite posizioni alternative per le finestre di dialogo installate per l'istanza di SPSS Statistics nella quale si sta copiando, è possibile copiare in una qualsiasi delle posizioni specificate: la finestra di dialogo verrà riconosciuta come finestra installata al successivo avvio di tale istanza.

Tipi di controlli

La tavolozza strumenti fornisce la serie di controlli che possono essere aggiunti a una finestra di dialogo personalizzata.

- **Elenco origine:** un elenco delle variabili sorgente dell'insieme di dati attivo. Per ulteriori informazioni, vedere Elenco origine a pag. 561.
- **Lista destinazione:** una destinazione per le variabili trasferite dall'elenco origine. Per ulteriori informazioni, vedere Lista destinazione a pag. 561.
- **Casella di controllo:** una casella di controllo singola. Per ulteriori informazioni, vedere Casella di controllo a pag. 563.
- **Casella combinata:** una casella combinata per la creazione di elenchi a discesa. Per ulteriori informazioni, vedere Casella combinata a pag. 564.
- **Controllo testo:** una casella di testo che accetta un testo arbitrario come input. Per ulteriori informazioni, vedere Controllo testo a pag. 565.
- **Controllo numerico:** una casella di testo limitata a valori numerici come input. Per ulteriori informazioni, vedere Controllo numerico a pag. 566.
- **Controllo testo statico:** un controllo per la visualizzazione di testo statico. Per ulteriori informazioni, vedere Controllo testo statico a pag. 566.
- **Gruppo di elementi:** un contenitore per raggruppare un insieme di controlli, ad esempio un insieme di caselle di controllo. Per ulteriori informazioni, vedere Gruppo di elementi a pag. 567.
- **Gruppo pulsanti di opzione:** un gruppo di pulsanti di opzione. Per ulteriori informazioni, vedere Gruppo pulsanti di opzione a pag. 567.
- **Gruppo di caselle di controllo:** un contenitore per un insieme di controlli abilitati o disabilitati in gruppo tramite un'unica casella di controllo. Per ulteriori informazioni, vedere Gruppo di caselle di controllo a pag. 568.
- **Browser di file:** un controllo per spostarsi nel file system per aprire o salvare un file. Per ulteriori informazioni, vedere Browser di file a pag. 569.
- **Pulsante sotto-finestra:** un pulsante per aprire una sotto-finestra. Per ulteriori informazioni, vedere Pulsante sotto-finestra a pag. 571.

Elenco origine

Il controllo Elenco di variabili sorgente visualizza l'elenco delle variabili dell'insieme di dati attivo disponibili all'utente finale della finestra di dialogo. È possibile visualizzare tutte le variabili dell'insieme di dati attivo (impostazione predefinita) oppure filtrare l'elenco in base al tipo e al livello di misurazione, ad esempio le variabili numeriche che hanno un livello di misurazione di scala. L'uso di un controllo Elenco origine implica l'uso di uno o più controlli Lista destinazione. Il controllo Elenco origine ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo.

Titolo. Un titolo facoltativo che viene visualizzato sopra il controllo. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo. Il testo specificato appare solo quando si posiziona il puntatore del mouse sull'area del titolo del controllo. Posizionando il puntatore su una delle variabili elencate si visualizza il nome e l'etichetta della variabile.

Tasto di scelta. Un carattere facoltativo da utilizzare come scelta rapida da tastiera nel titolo per il controllo. Il carattere appare sottolineato nel titolo. La scelta rapida viene attivata premendo Alt+[tasto di scelta rapida]. La proprietà Tasto di scelta non è supportata su Mac.

Trasferimento di variabili. Specifica se le variabili trasferite dall'elenco origine a una lista destinazione restano nell'elenco origine (Copia variabili) o vengono eliminati dall'elenco origine (Sposta variabili).

Filtro variabili. Consente di filtrare l'insieme di variabili visualizzate nel controllo. È possibile filtrare in base al tipo di variabili e al livello di misurazione, nonché specificare l'inclusione di insiemi a risposta multipla nell'elenco delle variabili. Fare clic sul pulsante con i puntini (...) per aprire la finestra di dialogo Filtro. Per ulteriori informazioni, vedere Applicazione di filtri agli elenchi di variabili a pag. 563.

Nota: il controllo Elenco origine non può essere aggiunto a una sotto-finestra.

Lista destinazione

Il controllo Lista destinazione fornisce una destinazione per le variabili trasferite dall'elenco origine. L'uso del controllo Lista destinazione presume la presenza di un controllo Elenco origine. È possibile specificare il trasferimento di una o di più variabili al controllo ed è possibile limitare i tipi di variabili che possono trasferite al controllo, ad esempio solo le variabili numeriche con un livello di misurazione nominale o ordinale. Il controllo Lista destinazione ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo. È l'identificatore da utilizzare per fare riferimento al controllo nel modello di sintassi.

Titolo. Un titolo facoltativo che viene visualizzato sopra il controllo. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo. Il testo specificato appare solo quando si posiziona il puntatore del mouse sull'area del titolo del controllo. Posizionando il puntatore su una delle variabili elencate si visualizza il nome e l'etichetta della variabile.

Tipo lista di destinazione. Specifica se è possibile trasferire un'unica variabile o più variabili al controllo.

Tasto di scelta. Un carattere facoltativo da utilizzare come scelta rapida da tastiera nel titolo per il controllo. Il carattere appare sottolineato nel titolo. La scelta rapida viene attivata premendo Alt+[tasto di scelta rapida]. La proprietà Tasto di scelta non è supportata su Mac.

Necessario per l'esecuzione. Specifica se è necessario un valore in questo controllo per il proseguimento dell'esecuzione. Se viene specificato Vero, i pulsanti OK e Incolla saranno disabilitati finché non viene specificato un valore per il controllo. Se viene specificato Falso, l'assenza di un valore nel controllo non ha alcun effetto sullo stato dei pulsanti OK e Incolla. L'impostazione predefinita è Vero.

Filtro variabili. Consente di limitare i tipi di variabili che possono essere trasferite al controllo. È possibile limitare in base al tipo di variabile e al livello di misurazione, nonché specificare se è possibile trasferire insiemi a riposta multipla al controllo. Fare clic sul pulsante con i puntini (...) per aprire la finestra di dialogo Filtro. Per ulteriori informazioni, vedere Applicazione di filtri agli elenchi di variabili a pag. 563.

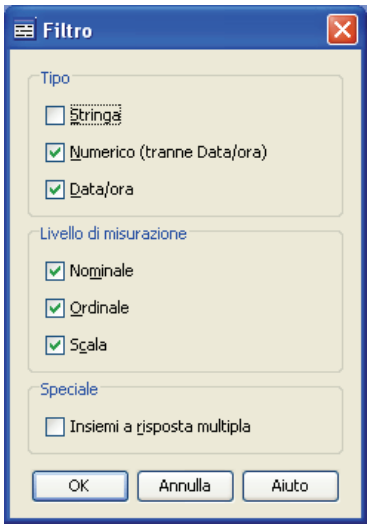
Sintassi. Specifica la sintassi dei comandi generata da questo controllo in fase di esecuzione e che può essere inserita nel modello di sintassi.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.
- Il valore `%%Questo valore%%` specifica il valore di runtime del controllo, ovvero l'elenco delle variabili trasferite al controllo. È l'impostazione predefinita.

Nota: il controllo Lista destinazione non può essere aggiunto a una sotto-finestra.

Applicazione di filtri agli elenchi di variabili

Figura 50-3
Finestra di dialogo Filtro



La finestra di dialogo Filtro, associata agli elenchi di variabili di origine e di destinazione, consente di filtrare i tipi di variabili dall'insieme di dati attivo che può essere visualizzato negli elenchi. È anche possibile specificare l'inclusione di insiemi a risposta multipla associati all'insieme di dati attivo. Le variabili numeriche comprendono tutti i formati numerici, tranne i formati data e ora.

Casella di controllo

Il controllo Casella di controllo è una semplice casella in grado di generare una sintassi dei comandi diversa per lo stato selezionato e deselezionato. Il controllo Casella di controllo ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo. È l'identificatore da utilizzare per fare riferimento al controllo nel modello di sintassi.

Titolo. L'etichetta visualizzata con la casella di controllo. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo.

Tasto di scelta. Un carattere facoltativo da utilizzare come scelta rapida da tastiera nel titolo per il controllo. Il carattere appare sottolineato nel titolo. La scelta rapida viene attivata premendo Alt+[tasto di scelta rapida]. La proprietà Tasto di scelta non è supportata su Mac.

Valore predefinito. Lo stato predefinito della casella di controllo (selezionata o deselezionata).

Sintassi verificata/non verificata. Specifica la sintassi dei comandi generata quando il controllo viene selezionato e quando viene deselezionato. Per includere la sintassi dei comandi nel modello di sintassi, utilizzare il valore della proprietà Identificatore. Sia la sintassi generata dalla proprietà Sintassi verificata che quella generata dalla proprietà Sintassi non verificata verrà inserita nella

posizione o nelle posizioni specificate dell'identificatore. Ad esempio, se l'identificatore è *casellacontrollo1*, in fase di esecuzione le istanze di `%%casellacontrollo1%%` nel modello di sintassi verranno sostituite dal valore della proprietà Sintassi verificata quando la casella è selezionata e dalla proprietà Sintassi non verificata quando la casella è deselezionata.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.

Casella combinata

Il controllo Casella combinata consente di creare un elenco a discesa in grado di generare una sintassi dei comandi specifica per la voce di elenco specificata. Il controllo Casella combinata ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo. È l'identificatore da utilizzare per fare riferimento al controllo nel modello di sintassi.

Titolo. Un titolo facoltativo che viene visualizzato sopra il controllo. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo.

Voci di elenco. Fare clic sul pulsante con i puntini (...) per aprire la finestra di dialogo Proprietà voci di elenco, che consente di specificare le voci di elenco del controllo casella combinata.

Tasto di scelta. Un carattere facoltativo da utilizzare come scelta rapida da tastiera nel titolo per il controllo. Il carattere appare sottolineato nel titolo. La scelta rapida viene attivata premendo Alt+[tasto di scelta rapida]. La proprietà Tasto di scelta non è supportata su Mac.

Sintassi. Specifica la sintassi dei comandi generata da questo controllo in fase di esecuzione e che può essere inserita nel modello di sintassi.

- Il valore `%%Questo valore%%` specifica il valore di runtime del controllo, ovvero il valore della proprietà Sintassi per la voce di elenco selezionata. È l'impostazione predefinita.
- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.

Specifica di voci di elenco per le caselle combinate

La finestra di dialogo Proprietà voci di elenco consente di specificare le voci di elenco di un controllo casella combinata.

Identificatore. Un identificatore univoco per la voce di elenco.

Nome. Il nome visualizzato nell'elenco a discesa per la voce di elenco. Il nome è un campo obbligatorio.

Impostazione predefinita. Specifica se la voce di elenco è la voce predefinita visualizzata nella casella combinata.

Sintassi. Specifica la sintassi dei comandi generata quando la voce di elenco viene selezionata.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.

È possibile aggiungere una nuova voce di elenco nella riga vuota alla base dell'elenco esistente. L'immissione di proprietà diverse dall'identificatore genera un identificatore univoco che è possibile mantenere o modificare. È possibile eliminare una voce di elenco facendo clic sulla cella *Identificatore* corrispondente alla voce e premendo Canc.

Controllo testo

Il controllo Testo è una semplice casella di testo in grado di accettare input arbitrario e ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo. È l'identificatore da utilizzare per fare riferimento al controllo nel modello di sintassi.

Titolo. Un titolo facoltativo che viene visualizzato sopra il controllo. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo.

Tasto di scelta. Un carattere facoltativo da utilizzare come scelta rapida da tastiera nel titolo per il controllo. Il carattere appare sottolineato nel titolo. La scelta rapida viene attivata premendo Alt+[tasto di scelta rapida]. La proprietà Tasto di scelta non è supportata su Mac.

Contenuto del testo. Specifica se il contenuto è arbitrario o se la casella di testo deve contenere una stringa che rispetti le regole sui nomi delle variabili di SPSS Statistics.

Valore predefinito. Il contenuto predefinito della casella di testo.

Necessario per l'esecuzione. Specifica se è necessario un valore in questo controllo per il proseguimento dell'esecuzione. Se viene specificato Vero, i pulsanti OK e Incolla saranno disabilitati finché non viene specificato un valore per il controllo. Se viene specificato Falso, l'assenza di un valore nel controllo non ha alcun effetto sullo stato dei pulsanti OK e Incolla. L'impostazione predefinita è Falso.

Sintassi. Specifica la sintassi dei comandi generata da questo controllo in fase di esecuzione e che può essere inserita nel modello di sintassi.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.
- Il valore `%%Questo valore%%` specifica il valore di runtime del controllo, ovvero il contenuto della casella di testo. È l'impostazione predefinita.

Controllo numerico

Il controllo Numerico è una casella di testo per l'immissione di un valore numerico e ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo. È l'identificatore da utilizzare per fare riferimento al controllo nel modello di sintassi.

Titolo. Un titolo facoltativo che viene visualizzato sopra il controllo. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo.

Tasto di scelta. Un carattere facoltativo da utilizzare come scelta rapida da tastiera nel titolo per il controllo. Il carattere appare sottolineato nel titolo. La scelta rapida viene attivata premendo Alt+[tasto di scelta rapida]. La proprietà Tasto di scelta non è supportata su Mac.

Tipo numerico. Specifica le eventuali limitazioni su ciò che può essere immesso. Il valore Numero reale specifica che non vi sono limitazioni sul valore immesso, se non che deve essere numerico. Il valore Numero intero specifica che il valore deve essere un numero intero.

Valore predefinito. Il valore predefinito, se presente.

Valore minimo. Il valore minimo consentito, se presente.

Valore massimo. Il valore massimo consentito, se presente.

Necessario per l'esecuzione. Specifica se è necessario un valore in questo controllo per il proseguimento dell'esecuzione. Se viene specificato Vero, i pulsanti OK e Incolla saranno disabilitati finché non viene specificato un valore per il controllo. Se viene specificato Falso, l'assenza di un valore nel controllo non ha alcun effetto sullo stato dei pulsanti OK e Incolla. L'impostazione predefinita è Falso.

Sintassi. Specifica la sintassi dei comandi generata da questo controllo in fase di esecuzione e che può essere inserita nel modello di sintassi.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.
- Il valore `%%Questo valore%%` specifica il valore di runtime del controllo, ovvero il valore numerico. È l'impostazione predefinita.

Controllo testo statico

Il controllo Testo statico consente di aggiungere un blocco di testo alla finestra di dialogo e ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo.

Titolo. Il contenuto del blocco di testo. Per contenuti su più righe, utilizzare `\n` per specificare interruzioni di riga.

Gruppo di elementi

Il controllo Gruppo di elementi è un contenitore di altri controlli, che consente di raggruppare e controllare la sintassi generata da più controlli. Ad esempio, si dispone di un insieme di caselle di controllo che specificano opzioni facoltative per un sottocomando, ma si desidera generare solo la sintassi per il sottocomando se è selezionata almeno una casella di controllo. A tale scopo è necessario utilizzare un controllo Gruppo di elementi come contenitore dei controlli della casella di controllo. Un Gruppo di elementi può contenere i tipi di controlli seguenti: casella di controllo, casella combinata, controllo di testo, controllo numerico, testo statico, gruppo pulsanti di opzione e browser di file. Il controllo Gruppo di elementi ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo. È l'identificatore da utilizzare per fare riferimento al controllo nel modello di sintassi.

Titolo. Un titolo facoltativo per il gruppo. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Necessario per l'esecuzione. Vero specifica che i pulsanti OK e Incolla saranno disabilitati finché almeno un controllo del gruppo non genera la sintassi dei comandi. Se viene specificato Falso, l'assenza della sintassi di runtime da parte di questo controllo non ha alcun effetto sullo stato dei pulsanti OK e Incolla. L'impostazione predefinita è Falso.

Ad esempio, il gruppo è composto da un insieme di caselle di controllo, ciascuna delle quali genera solo la sintassi nello stato selezionato. Se Necessario per l'esecuzione è impostato su Vero e tutte le caselle sono deselezionate, OK e Incolla saranno disabilitati.

Sintassi. Specifica la sintassi dei comandi generata da questo controllo in fase di esecuzione e che può essere inserita nel modello di sintassi.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.
- È possibile includere identificatori per tutti i controlli contenuti nel gruppo di elementi. In fase di esecuzione gli identificatori vengono sostituiti con la sintassi generata dai controlli.
- Il valore `%%QuestoValore%%` genera un elenco delimitato da spazi vuoti della sintassi generata da ogni controllo del gruppo di elementi, nello stesso ordine in cui appaiono all'interno del gruppo (dall'alto in basso). È l'impostazione predefinita. Se `%%QuestoValore%%` è specificato e non viene generata alcuna sintassi dai controlli del gruppo di elementi, quest'ultimo non genera alcuna sintassi dei comandi.

Gruppo pulsanti di opzione

Il controllo Gruppo pulsanti di opzione è un contenitore per un insieme di pulsanti di opzione e ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo. È l'identificatore da utilizzare per fare riferimento al controllo nel modello di sintassi.

Titolo. Un titolo facoltativo per il gruppo. Se omesso, il bordo del gruppo non viene visualizzato. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo.

Pulsanti di opzione. Fare clic sul pulsante con i puntini (...) per aprire la finestra di dialogo Proprietà gruppo pulsanti di opzione, che consente di specificare le proprietà dei pulsanti di opzione, nonché di aggiungere o eliminare pulsanti dal gruppo.

Sintassi. Specifica la sintassi dei comandi generata da questo controllo in fase di esecuzione e che può essere inserita nel modello di sintassi.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.
- Il valore `%%Questo valore%%` specifica il valore di runtime del gruppo di pulsanti di opzione, ovvero il valore della proprietà Sintassi per il pulsante di opzione selezionato. È l'impostazione predefinita.

Definizione di pulsanti di opzione

La finestra di dialogo Proprietà gruppo pulsanti di opzione consente di specificare un gruppo di pulsanti di opzione.

Identificatore. Un identificatore univoco per il pulsante di opzione.

Nome. Il nome visualizzato accanto al pulsante di opzione. Il nome è un campo obbligatorio.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo.

Tasto di scelta. Un carattere facoltativo nel nome da utilizzare come tasto di scelta. Il carattere specificato deve essere presente nel nome.

Impostazione predefinita. Specifica se il pulsante di opzione è la selezione predefinita.

Sintassi. Specifica la sintassi dei comandi generata quando il pulsante di opzione viene selezionato.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.

È possibile aggiungere un nuovo pulsante di opzione nella riga vuota alla base dell'elenco esistente. L'immissione di proprietà diverse dall'identificatore genera un identificatore univoco che è possibile mantenere o modificare. È possibile eliminare un pulsante di opzione facendo clic sulla cella *Identificatore* corrispondente al pulsante e premendo Canc.

Gruppo di caselle di controllo

Il controllo Gruppo di caselle di controllo è un contenitore per un insieme di controlli abilitati o disabilitati in gruppo tramite un'unica casella di controllo. Un Gruppo di caselle di controllo può contenere i tipi di controlli seguenti: casella di controllo, casella combinata, controllo di testo, controllo numerico, testo statico, gruppo pulsanti di opzione e browser di file. I controllo Gruppo di caselle di controllo ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo. È l'identificatore da utilizzare per fare riferimento al controllo nel modello di sintassi.

Titolo. Un titolo facoltativo per il gruppo. Se omesso, il bordo del gruppo non viene visualizzato. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Titolo casella di controllo. Un'etichetta facoltativa visualizzata con la casella di controllo principale. Supporta `\n` per specificare le interruzioni di riga.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo.

Tasto di scelta. Un carattere facoltativo da utilizzare come scelta rapida da tastiera nel titolo per il controllo. Il carattere appare sottolineato nel titolo. La scelta rapida viene attivata premendo Alt+[tasto di scelta rapida]. La proprietà Tasto di scelta non è supportata su Mac.

Valore predefinito. Lo stato predefinito della casella di controllo principale (selezionata o deselezionata).

Sintassi verificata/non verificata. Specifica la sintassi dei comandi generata quando il controllo viene selezionato e quando viene deselezionato. Per includere la sintassi dei comandi nel modello di sintassi, utilizzare il valore della proprietà Identificatore. Sia la sintassi generata dalla proprietà Sintassi verificata che quella generata dalla proprietà Sintassi non verificata verrà inserita nella posizione o nelle posizioni specificate dell'identificatore. Ad esempio, se l'identificatore è *gruppocasellecontrollo1*, in fase di esecuzione le istanze di `%%gruppocasellecontrollo1%%` nel modello di sintassi verranno sostituite dal valore della proprietà Sintassi verificata quando la casella è selezionata e dalla proprietà Sintassi non verificata quando la casella è deselezionata.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.
- È possibile includere identificatori per tutti i controlli contenuti nel gruppo di caselle di controllo. In fase di esecuzione gli identificatori vengono sostituiti con la sintassi generata dai controlli.
- Il valore `%%QuestoValore%%` può essere utilizzato sia nella proprietà Sintassi verificata che nella proprietà Sintassi non verificata. Genera un elenco delimitato da spazi vuoti della sintassi generata da ogni controllo del gruppo di caselle di controllo, nello stesso ordine in cui appaiono all'interno del gruppo (dall'alto in basso).
- Per impostazione predefinita, la proprietà Sintassi verificata ha il valore `%%QuestoValore%%` e la proprietà Sintassi non verificata è vuota.

Browser di file

Il controllo Browser di file è composto da una casella di testo per un percorso di file e un pulsante Sfoglia che apre una finestra di dialogo standard di SPSS Statistics per aprire o salvare un file. Il controllo Browser di file ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo. È l'identificatore da utilizzare per fare riferimento al controllo nel modello di sintassi.

Titolo. Un titolo facoltativo che viene visualizzato sopra il controllo. Per titoli su più righe, utilizzare `\n` per specificare interruzioni di riga.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo.

Tasto di scelta. Un carattere facoltativo da utilizzare come scelta rapida da tastiera nel titolo per il controllo. Il carattere appare sottolineato nel titolo. La scelta rapida viene attivata premendo Alt+[tasto di scelta rapida]. La proprietà Tasto di scelta non è supportata su Mac.

Operazione su file system. Specifica se la finestra di dialogo avviata dal pulsante Sfoglia è adatta per aprire o salvare i file. Un valore Apri indica che la finestra di dialogo Sfoglia convalida l'esistenza del file specificato. Un valore Salva indica che la finestra di dialogo Sfoglia non convalida l'esistenza del file specificato.

Tipo di browser. Specifica se la finestra di dialogo Sfoglia viene utilizzata per selezionare un file (Trova file) o per selezionare una cartella (Trova cartella).

Filtro dei file. Fare clic sul pulsante con i puntini (...) per aprire la finestra di dialogo Filtro dei file, che consente di specificare i tipi di file disponibili per la finestra di dialogo Salva o Apri. Per impostazione predefinita sono consentiti tutti i tipi di file.

Tipo di file system. Nella modalità di analisi distribuita, specifica se la finestra di dialogo deve cercare nel file system sul quale è in esecuzione SPSS Statistics Server o nel file system del computer locale: Server indica di cercare nel file system del server, mentre Client indica di cercare nel file system del computer locale. La proprietà non ha alcun effetto in modalità di analisi locale.

Necessario per l'esecuzione. Specifica se è necessario un valore in questo controllo per il proseguimento dell'esecuzione. Se viene specificato Vero, i pulsanti OK e Incolla saranno disabilitati finché non viene specificato un valore per il controllo. Se viene specificato Falso, l'assenza di un valore nel controllo non ha alcun effetto sullo stato dei pulsanti OK e Incolla. L'impostazione predefinita è Falso.

Sintassi. Specifica la sintassi dei comandi generata da questo controllo in fase di esecuzione e che può essere inserita nel modello di sintassi.

- È possibile specificare qualsiasi sintassi dei comandi valida, nonché utilizzare `\n` per le interruzioni di riga.
- Il valore `%%QuestoValore%%` specifica il valore di runtime della casella di testo, ovvero il percorso del file, specificato manualmente o popolato dalla finestra di dialogo Sfoglia. È l'impostazione predefinita.

Filtro dei tipi di file

Figura 50-4
Finestra di dialogo Filtro dei file

La finestra di dialogo Filtro dei file consente di specificare i tipi di file visualizzati negli elenchi a discesa Tipo file e Salva con nome per le finestre di dialogo Apri e Salva aperte da un controllo Browser di file system. Per impostazione predefinita sono consentiti tutti i tipi di file.

Per specificare i tipi di file non elencati esplicitamente nella finestra di dialogo:

- Selezionare Altro.
- Digitare un nome per il tipo di file.
- Immettere un tipo di file in forma *.suffisso, ad esempio *.xls. È possibile specificare più tipi di file, ciascuno delimitato da un punto e virgola.

Pulsante sotto-finestra

Il controllo Pulsante sotto-finestra specifica un pulsante per aprire una sotto-finestra e fornisce l'accesso al Generatore di finestre di dialogo per la sotto-finestra. Il Pulsante sotto-finestra ha le proprietà seguenti:

Identificatore. L'identificatore univoco del controllo.

Titolo. Il testo visualizzato nel pulsante.

Descrizioni. Un testo Suggerimento facoltativo che viene visualizzato quando l'utente posiziona il puntatore del mouse sopra il controllo.

Sotto-finestra. Fare clic sul pulsante con i puntini (...) per aprire il Generatore di finestre di dialogo personalizzate per la sotto-finestra. Per aprire il generatore è anche possibile fare doppio clic sul pulsante Sotto-finestra.

Tasto di scelta. Un carattere facoltativo da utilizzare come scelta rapida da tastiera nel titolo per il controllo. Il carattere appare sottolineato nel titolo. La scelta rapida viene attivata premendo Alt+[tasto di scelta rapida]. La proprietà Tasto di scelta non è supportata su Mac.

Nota: il controllo Pulsante sotto-finestra non può essere aggiunto a una sotto-finestra.

Proprietà di una sotto-finestra

Per visualizzare e impostare le proprietà di una sotto-finestra:

- Aprire la sotto-finestra facendo doppio clic sul pulsante corrispondente alla sotto-finestra nella finestra di dialogo principale, oppure fare clic sul pulsante della sotto-finestra e sul pulsante con i puntini (...) corrispondente alla proprietà Sotto-finestra.
- Nella sotto-finestra, fare clic sull'area di disegno al di fuori dei controlli. Senza alcun controllo sull'area di disegno, le proprietà di una sotto-finestra sono sempre visibili.

Nome sotto-finestra. L'identificatore univoco della sotto-finestra. La proprietà Nome sotto-finestra è obbligatoria.

Nota: se si specifica il Nome sotto-finestra come identificatore nel modello di sintassi, ad esempio `%%Nome sotto-finestra%%`, verrà sostituito in fase di esecuzione con un elenco delimitato da spazi vuoti della sintassi generata da ogni controllo della sotto-finestra, nello stesso ordine in cui vengono visualizzati (dall'alto in basso e da sinistra a destra).

Titolo. Specifica il testo da visualizzare nella barra del titolo della sotto-finestra. La proprietà Titolo è facoltativa ma consigliata.

File di Aiuto. Specifica il percorso a un file di Aiuto facoltativo per la sotto-finestra. È il file che verrà avviato quando l'utente fa clic sul pulsante Aiuto nella sotto-finestra e può essere lo stesso file di Aiuto specificato per la finestra di dialogo principale. I file di Aiuto devono essere in formato HTML. Per ulteriori informazioni, vedere la descrizione della proprietà File di Aiuto per Proprietà finestra di dialogo.

Sintassi. Fare clic sul pulsante con i puntini (...) per aprire il Modello di sintassi. Per ulteriori informazioni, vedere Creazione del modello di sintassi a pag. 555.

Finestre di dialogo personalizzate per i comandi di estensione

I comandi di estensione sono comandi di SPSS Statistics definiti dall'utente e implementati nel linguaggio di programmazione Python o R. Una volta distribuito a un'istanza di SPSS Statistics, un comando di estensione viene eseguito allo stesso modo di qualsiasi comando incorporato di SPSS Statistics.

È possibile utilizzare il Generatore di finestre di dialogo personalizzate per creare finestre di dialogo per i comandi di estensione, nonché installare finestre di dialogo per comandi di estensione creati da altri utenti.

Creazione di finestre di dialogo personalizzate per i comandi di estensione

Indipendentemente dall'utente che ha creato un comando di estensione, è possibile creare una finestra di dialogo personalizzata appositamente per tale comando. Il modello di sintassi per la finestra di dialogo deve generare la sintassi per il comando di estensione. Se la finestra di dialogo è solo per uso personale, installarla. Se che il comando di estensione è già distribuito sul proprio sistema, sarà possibile eseguire il comando dalla finestra di dialogo installata.

Se si sta creando una finestra di dialogo personalizzata per un comando di estensione e si desidera condividerla con altri utenti, è necessario salvare le specifiche della finestra in un file pacchetto finestra di dialogo personalizzato (*.spd*) e quindi distribuire tale file. Il file pacchetto finestra di dialogo personalizzato non conterrà le specifiche del comando di estensione stesso, ma solo quelle della finestra di dialogo personalizzata. Sarà necessario fornire separatamente agli utenti finali il file XML che specifica la sintassi del comando di estensione e il codice di implementazione scritto in Python o R.

Installazione di comandi di estensione con finestre di dialogo personalizzate associate

Un comando di estensione con una finestra di dialogo personalizzata associata è composto da tre elementi: un file XML che specifica la sintassi del comando, un file di codice (un modulo Python o un pacchetto R) che implementa il comando e un file pacchetto finestra di dialogo personalizzato che contiene le specifiche della finestra di dialogo personalizzata.

File pacchetto finestra di dialogo personalizzato. Installare il file pacchetto finestra di dialogo personalizzato dalla voce Installa finestra di dialogo del menu Strumenti.

File XML di specifica della sintassi e codice di implementazione. Il file XML che specifica la sintassi del comando di estensione e il codice di implementazione (modulo Python o pacchetto R) devono essere inseriti nella cartella *extensions* sotto la directory di installazione di SPSS Statistics. Per Mac, la directory di installazione si riferisce alla directory *Contents* nel bundle dell'applicazione per SPSS Statistics.

- Se non si dispone di autorizzazioni di scrittura sulla directory di installazione di SPSS Statistics o se si desidera memorizzare il file XML e il codice di implementazione in un'altra posizione, è possibile specificare una o più posizioni alternative definendo la variabile di ambiente *SPSS_EXTENSIONS_PATH*. Se presenti, i percorsi specificati in *SPSS_EXTENSIONS_PATH* hanno la precedenza sulla directory di installazione di SPSS Statistics. Per posizioni multiple, separarle con il punto e virgola su Windows e con i due punti su Linux e Mac.
- Per un comando di estensione implementato in Python, è sempre possibile memorizzare il modulo Python associato in una posizione sul percorso di ricerca di Python, ad esempio la directory *site-packages* di Python.
- Per un comando di estensione implementato in R, il pacchetto R contenente il codice di implementazione deve essere installato nella directory contenente il file di specifica della sintassi XML. In alternativa è sempre possibile installarlo nella posizione predefinita della piattaforma associata, ad esempio *R_Home/library* su Windows, dove *R_Home* è la posizione di installazione di R e *library* è una sottocartella di tale directory. Per sapere come installare i pacchetti R, consultare la guida di installazione e amministrazione distribuita con R.

Per informazioni sulla creazione di comandi di estensione, vedere l'articolo *"Writing SPSS Statistics Extension Commands"*, disponibile sul sito Developer Central all'indirizzo *http://www.spss.com/devcentral* . È consigliabile anche consultare il capitolo relativo ai comandi di estensione in *Programming and Data Management for SPSS Statistics*, disponibile in formato PDF nel sito *http://www.spss.com/spss/data_management_book.htm*.

Creazione di versioni localizzate di finestre di dialogo personalizzate

È possibile creare versioni localizzate delle finestre di dialogo personalizzate per tutte le lingue supportate da SPSS Statistics. È possibile localizzare tutte le stringhe visualizzate in una finestra di dialogo personalizzata, nonché il file di Aiuto facoltativo.

Per localizzare le stringhe delle finestre di dialogo

- Creare una copia del file delle proprietà associato alla finestra di dialogo. Il file delle proprietà contiene tutte le stringhe localizzabili associate alla finestra di dialogo. Per impostazione predefinita, si trova nella cartella *ext/lib/<Finestra di dialogo>* della directory di installazione di SPSS Statistics, che si trova nella directory principale della directory di installazione per Windows e Linux e nella cartella*Contents/bin* nel bundle dell'applicazione di SPSS Statistics 17.0 per Mac. La copia deve risiedere nella stessa cartella e non in una sottocartella.

  Se sono state specificate posizioni alternative per le finestre di dialogo installate (utilizzando la variabile di ambiente *SPSS_CDIALOGS_PATH*), la copia deve risiedere nella cartella *<Finestra di dialogo>* nella posizione alternativa corrispondente. Per ulteriori informazioni, vedere Gestione delle finestre di dialogo personalizzate a pag. 558.

- Rinominare la copia *<Finestra di dialogo>_<identificatore lingua>.properties*, utilizzando gli identificatori di lingua elencati nella tabella sottostante. Ad esempio, se il nome della finestra di dialogo è *finestradialogo* e si desidera crearne una versione giapponese, il file delle proprietà localizzato deve essere chiamato *finestradialogo_ja.properties*. I file delle proprietà localizzati devono inoltre essere aggiunti manualmente ai file pacchetto personalizzati creati per la finestra di dialogo.

- Aprire il nuovo file delle proprietà con un editor di testo che supporta UTF-8, ad esempio Notepad su Windows o l'applicazione TextEdit su Mac. Modificare i valori associati alle proprietà che devono essere localizzate, ma non modificare i nomi delle proprietà. Le proprietà associate a un controllo specifico sono precedute dall'identificatore del controllo. Ad esempio, la proprietà Suggerimento per un controllo con l'identificatore *pulsante_opzioni* è *pulsante_opzioni_suggerimento_ETICHETTA*. Le proprietà del titolo sono chiamate semplicemente *<identificatore>_ETICHETTA*, ad esempio *pulsante_opzioni_ETICHETTA*. I due punti utilizzati nei valori delle proprietà vanno inseriti tra caratteri di escape, ad esempio `\:`.

  Quando la finestra di dialogo viene avviata, SPSS Statistics cercherà un file delle proprietà il cui identificatore di lingua corrisponda alla lingua corrente, specificata dall'elenco a discesa Lingua nella scheda Generale della finestra di dialogo Opzioni. Se tale file delle proprietà non viene trovato, verrà utilizzato il file predefinito *<Finestra di dialogo>.properties*.

Per localizzare il file di Aiuto

- Creare una copia del file di Aiuto associato alla finestra di dialogo e localizzare il testo nella lingua desiderata. La copia deve risiedere nella stessa cartella del file di Aiuto e non in una sottocartella. Il file di Aiuto dovrebbe trovarsi nella cartella *ext/lib/<Finestra di dialogo>* della directory di installazione di SPSS Statistics, che si trova nella directory principale della directory di installazione per Windows e Linux e nella cartella *Contents/bin* nel bundle dell'applicazione di SPSS Statistics 17.0 per Mac.

Se sono state specificate posizioni alternative per le finestre di dialogo installate (utilizzando la variabile di ambiente *SPSS_CDIALOGS_PATH*), la copia deve risiedere nella cartella *<Finestra di dialogo>* nella posizione alternativa corrispondente. Per ulteriori informazioni, vedere Gestione delle finestre di dialogo personalizzate a pag. 558.

- Rinominare la copia *<File di Aiuto>_<identificatore lingua>*, utilizzando gli identificatori di lingua elencati nella tabella sottostante. Ad esempio, se il file di Aiuto è *aiuto.htm* e si desidera creare una versione in tedesco del file, il file di Aiuto localizzato deve essere chiamato *aiuto_de.htm*. I file di Aiuto localizzati devono inoltre essere aggiunti manualmente ai file pacchetto personalizzati creati per la finestra di dialogo.

  In presenza di file supplementari quali i file di immagine, che devono a loro volta essere localizzati, sarà necessario modificare manualmente i percorsi corrispondenti nel file di Aiuto principale in modo che puntino alle versioni localizzate, che devono essere memorizzate con le versioni originali.

  Quando la finestra di dialogo viene avviata, SPSS Statistics cercherà un file di Aiuto il cui identificatore di lingua corrisponda alla lingua corrente, specificata dall'elenco a discesa Lingua nella scheda Generale della finestra di dialogo Opzioni. Se tale file di Aiuto non viene trovato, verrà utilizzato il file di Aiuto specificato per la finestra di dialogo (specificato nella proprietà File di Aiuto di Proprietà finestra di dialogo).

Identificatori lingua

de	Tedesco
en	Inglese
es	Spagnolo
fr	Francese
it	Italiano
ja	Giapponese
ko	Coreano
pl	Polacco
ru	Russo
zh_cn	Cinese semplificato
zh_tw	Cinese tradizionale

Nota: il testo nelle finestre di dialogo personalizzate e nei file di Aiuto associati non è limitato alle lingue supportate da SPSS Statistics. È possibile scrivere la finestra di dialogo e il testo dell'Aiuto in qualsiasi lingua senza necessità di creare proprietà e file di Aiuto specifici per la lingua. Tutti gli utenti della finestra di dialogo vedranno quindi il testo in tale lingua.

Capitolo
51

Sessioni Utilità SPSS Production

Le sessioni Utilità SPSS Production consentono di eseguire SPSS Statistics in modo automatico. Il programma viene eseguito in modo automatico e termina dopo l'esecuzione dell'ultimo comando. In tal modo, si possono svolgere contemporaneamente altre attività o si può programmare l'esecuzione automatica della sessione agli orari specificati. Le sessioni Utilità SPSS Production sono utili in caso di esecuzione ripetuta degli stessi insiemi di analisi che richiedono tempi di elaborazione piuttosto lunghi, ad esempio i report settimanali.

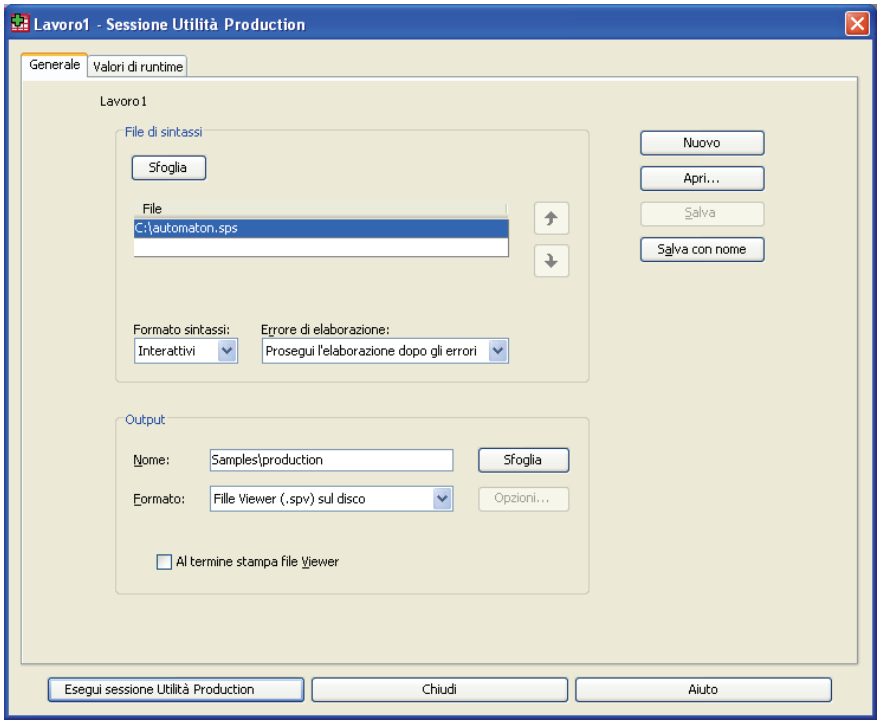
Le sessioni utilizzano i file della sintassi dei comandi per indicare a SPSS Statistics le operazioni da eseguire. Un file della sintassi dei comandi è un semplice file di testo che contiene la sintassi dei comandi. Per creare il file è possibile utilizzare qualsiasi editor di testo. È inoltre possibile generare la sintassi dei comandi incollando le selezioni della finestra di dialogo in una finestra di sintassi o modificando il file giornale. Per ulteriori informazioni, vedere Utilizzo della sintassi dei comandi in Capitolo 13 a pag. 268.

Per creare una sessione Utilità SPSS Production:

- Dai menu di qualsiasi finestra, scegliere:

  Strumenti
  Sessione Utilità SPSS Production

Figura 51-1
Finestra di dialogo Sessione Utilità SPSS Production



▶ Fare clic su Nuovo per creare una nuova sessione Utilità SPSS Production o su Apri per aprirne una esistente.

Nota: I file Production Facility (.spp) creati con le versioni precedenti alla 16.0 non possono essere utilizzati con la versione 16.0 o con le versioni successive. È disponibile un'utilità per convertire i file Production Facility di Windows e Macintosh in file SPSS Production (.spj). Per ulteriori informazioni, vedere Conversione dei file della sessione Utilità SPSS Production a pag. 584.

▶ Selezionare uno o più file della sintassi dei comandi.

Nota: nella modalità di analisi distribuita, è necessario immettere manualmente le posizioni dei file della sintassi e i percorsi relativi fanno riferimento al server remoto. Per ulteriori informazioni, vedere Specifiche di percorsi assoluti e relativi in Capitolo 4 a pag. 75.

▶ Selezionare il nome, la posizione e il formato del file di output.

▶ Fare clic su Salva o su Salva con nome per salvare la sessione Utilità SPSS Production.

Formato sintassi. Controlla il formato delle regole della sintassi utilizzate per la sessione.

- **Interattive.** Ciascun comando deve terminare con un punto. I punti possono essere inseriti in qualsiasi punto del comando. Sebbene i comandi possano continuare su più righe, l'uso di un punto come ultimo carattere non vuoto viene interpretato come la fine del comando. Le righe di continuazione ed i nuovi comandi possono iniziare in qualsiasi punto della nuova

riga. Queste regole "interattive" si applicano quando si selezionano e si eseguono comandi in una finestra di sintassi.

- **Batch.** Ciascun comando deve iniziare all'inizio di una nuova riga (non devono essere inseriti spazi vuoti prima dell'inizio del comando), mentre le linee di continuazione devono essere rientrate di almeno uno spazio. Per rientrare i nuovi comandi, è possibile utilizzare il segno più ("+"), la barra o un punto come primo carattere all'inizio della riga, quindi rientrare il comando vero e proprio. Il punto alla fine del comando è facoltativo. Questa impostazione è compatibile con le regole della sintassi dei file dei comandi inclusi con il comando `INCLUDE`.

  Nota: *non* utilizzare l'opzione Batch se i file di sintassi contengono la sintassi del comando `GGRAPH` che include le istruzioni `GPL`. Le istruzioni `GPL` verranno eseguite soltanto con le regole interattive.

Errore di elaborazione. Controlla la modalità con cui vengono gestite le condizioni di errore della sessione.

- **Prosegui l'elaborazione dopo gli errori.** Gli errori della sessione non interrompono automaticamente l'elaborazione dei comandi. I comandi dei file della sessione vengono considerati parte del normale flusso dei comandi, quindi l'elaborazione dei comandi viene continuata normalmente.
- **Arresta immediatamente l'elaborazione.** L'elaborazione dei comandi si interrompe appena si verifica il primo errore nella sessione. Ciò è compatibile con il comportamento dei file dei comandi inclusi con il comando `INCLUDE`.

Output. Controlla il nome, la posizione e il formato dei risultati della sessione Utilità SPSS Production. Sono disponibili le seguenti opzioni di formato:

- **File Viewer (.spv) sul disco.** I risultati vengono salvati nel formato Viewer SPSS Statistics nella posizione file specificata.
- **File Viewer (.spv) sul repository PES.** Richiede l'opzione Adaptor for Predictive Enterprise Services.
- **Report Web (.spw) in repository PES.** Richiede l'opzione Adaptor for Predictive Enterprise Services.
- **Word/RTF (*.doc).** Le tabelle pivot vengono esportate come tabelle Word con tutti gli attributi di formattazione; ad esempio bordi delle celle, stili del carattere e colori dello sfondo. L'output testuale viene esportato come HTML formattato. I grafici, i diagrammi ad albero e le viste del modello vengono inclusi nel formato PNG.

  Nota: alcune tabelle molto ampie non vengono visualizzate correttamente in Microsoft Word.
- **Excel (*.xls).** Le righe, le colonne e le celle delle tabelle pivot vengono esportate come righe, colonne e celle di Excel, con tutti gli attributi di formattazione, ad esempio, con i bordi delle celle, gli stili del carattere, il colore di sfondo delle celle e così via. L'output testuale viene esportato con tutti gli attributi del carattere. Ogni riga nell'output testuale è una riga del file di Excel, con tutti i contenuti di una riga raccolti in una singola cella. I grafici, i diagrammi ad albero e le viste del modello vengono inclusi nel formato PNG.
- **HTML (*.htm).** Le tabelle pivot vengono esportate come tabelle HTML. L'output testuale viene esportato come HTML preformattato. I grafici, i diagrammi ad albero e le viste dei modelli vengono incorporati per riferimento e possono essere esportati in un formato adatto all'inserimento in documenti HTML (ad esempio, PNG e JPEG).

- **Portable Document Format (*.pdf).** Tutto l'output viene esportato così come appare in Anteprima di stampa con tutti gli attributi di formattazione invariati.
- **File PowerPoint (*.ppt).** Le tabelle pivot vengono esportate come tabelle Word ed incorporate in diapositive diverse del file PowerPoint; una diapositiva per ciascuna tabella pivot. Tutti gli attributi di formattazione della tabella pivot (quali i bordi delle celle, gli stili dei caratteri, i colori di sfondo, ecc.) vengono mantenuti. I grafici, i diagrammi ad albero e le viste del modello vengono esportati in formato TIFF. L'output testuale non è incluso.

  L'esportazione in PowerPoint è disponibile solo nei sistemi operativi Windows.
- **Testo (*.txt).** I formati di output testuale includono testo normale, UTF-8 e UTF-16. Le tabelle pivot possono essere esportate in formato delimitato da tabulazioni o da spazi. Tutto l'output testuale viene esportato in formato delimitato da spazi. Per i grafici, i diagrammi ad albero e le viste del modello viene inserita una riga nel file di testo per ogni immagine, che indica il nome del file di immagine.

Al termine stampa file Viewer. Invia il file di output finale del Viewer alla stampante al termine della sessione Utilità SPSS Production.

Esegui sessione Utilità SPSS Production. Esegue la sessione Utilità SPSS Production separatamente. Questa opzione è particolarmente utile per testare le nuove sessioni Utilità SPSS Production prima dell'implementazione.

Sessioni Utilità SPSS Production con i comandi OUTPUT

Le sessioni dell'utilità SPSS Production supportano i comandi `OUTPUT`, come ad esempio `OUTPUT SAVE`, `OUTPUT ACTIVATE` e `OUTPUT NEW`. I comandi `OUTPUT SAVE` eseguiti nel corso di una sessione Utilità SPSS Production scriveranno il contenuto dei documenti di output specificati nelle posizioni specificate, in aggiunta al file di output creato dalla sessione Utilità SPSS Production. Quando si utilizza `OUTPUT NEW` per creare un nuovo documento di output, è consigliabile eseguire esplicitamente il salvataggio con il comando `OUTPUT SAVE`.

Un file di output della sessione Utilità SPSS Production è costituito dal contenuto del documento di output attivo alla fine della sessione. Per le sessioni contenenti i comandi `OUTPUT`, il file di output può non contenere tutto l'output creato nella sessione. Ad esempio, si supponga che la sessione dell'utilità SPSS Production sia costituita da alcune procedure seguite da un comando `OUTPUT NEW`, seguito da più procedure ma senza comandi `OUTPUT`. Il comando `OUTPUT NEW` definisce un nuovo documento di ouput attivo. Al termine della sessione Utilità SPSS Production, conterrà l'output generato soltanto dalle procedure eseguite dopo il comando `OUTPUT NEW`.

Opzioni HTML

Opzioni tabella. Per il formato HTML non sono disponibili opzioni relative alla tabella. Tutte le tabelle pivot vengono convertite in tabelle HTML.

Opzioni immagine. I tipi di formato immagine disponibili sono: EPS, JPEG, TIFF, PNG e BMP. Nei sistemi operativi Windows è disponibile anche il formato EMF (Metafile avanzato). È possibile anche scalare le dimensioni dell'immagine nell'intervallo 1% - 200%.

Opzioni di PowerPoint

Opzioni tabella. È possibile utilizzare le voci di output di Viewer come titoli di diapositiva. Ciascuna diapositiva contiene una singola voce di output. Il titolo è costituito dalla voce corrispondente all'elemento desiderato nella vista riassuntiva di Viewer.

Opzioni immagine. È possibile scalare le dimensioni dell'immagine nell'intervallo 1% - 200%. Tutte le immagini vengono esportate nel formato TIFF di PowerPoint.

Nota: il formato PowerPoint è disponibile solo nei sistemi operativi Windows e richiede PowerPoint 97 o versione successiva.

Opzioni PDF

Segnalibri incorporati. Questa opzione include nel documento PDF i segnalibri corrispondenti alle voci di Vista riassuntiva del Viewer. Analogamente alla Vista riassuntiva del Viewer, i segnalibri facilitano lo spostamento nei documenti contenenti un elevato numero di oggetti di output.

Caratteri incorporati. L'incorporazione dei caratteri garantisce che il documento PDF avrà lo stesso aspetto su tutti i computer. Altrimenti, se alcuni caratteri del documento non fossero disponibili sul computer usato per visualizzare (o stampare) il documento PDF, la sostituzione dei caratteri potrebbe generare risultati deludenti.

Opzioni testo

Opzioni tabella. È possibile esportare le tabelle pivot in formato delimitato da tabulazioni o da spazi. Per il formato delimitato da spazi è anche possibile controllare:

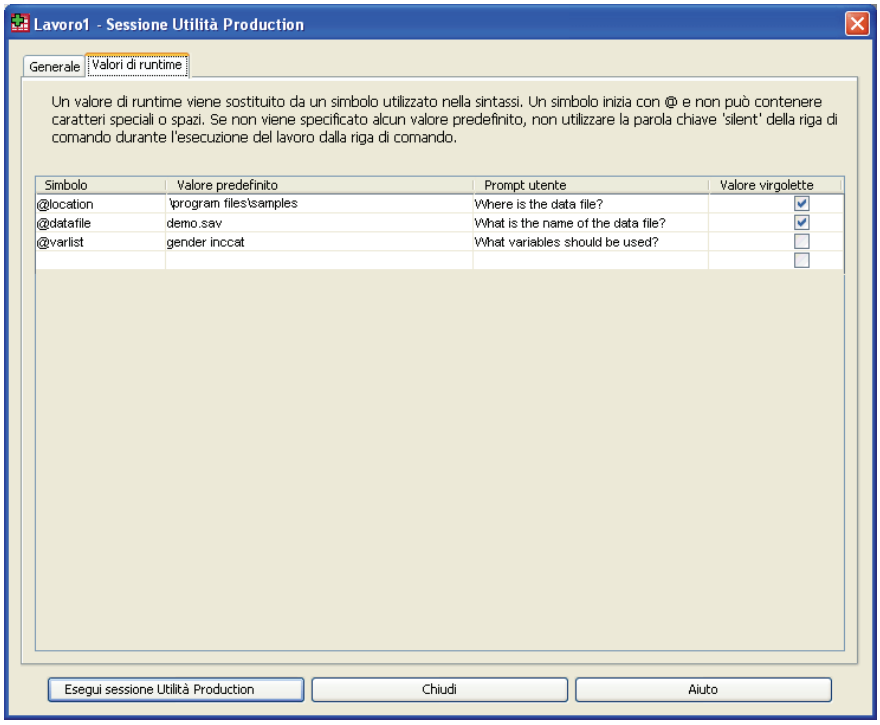
- **Larghezza colonna.** L'opzione Adatta non manda a capo il testo dei contenuti delle colonne e di conseguenza l'ampiezza di ciascuna colonna è pari a quella dell'etichetta o del valore più ampio contenuto in tale colonna. L'opzione Personalizzata imposta una larghezza massima di colonna che viene applicata a tutte le colonne della tabella; i valori che superano tale larghezza vengono mandati a capo nella riga successiva di tale colonna.
- **Carattere del bordo riga/colonna.** Queste opzioni consentono di controllare i caratteri utilizzati per creare bordi di riga e di colonna. Per disattivare la visualizzazione dei bordi di riga e di colonna, immettere spazi vuoti al posto dei valori.

Opzioni immagine. I tipi di formato immagine disponibili sono: EPS, JPEG, TIFF, PNG e BMP. Nei sistemi operativi Windows è disponibile anche il formato EMF (Metafile avanzato). È possibile anche scalare le dimensioni dell'immagine nell'intervallo 1% - 200%.

Valori di runtime

I valori di runtime definiti in un file di sessione Utilità SPSS Production e utilizzati in un file di sintassi dei comandi consentono di semplificare operazioni come l'esecuzione della stessa analisi per file di dati diversi o l'esecuzione di una serie di comandi per diversi insiemi di variabili. Ad esempio, è possibile definire il valore di runtime *@datafile* in modo che venga richiesto il nome di un file dati ad ogni esecuzione di una sessione Utilità SPSS Production in cui venga utilizzata la stringa *@datafile* al posto di un nome di file nel file della sintassi del comando.

Figura 51-2
Scheda Valori di runtime della finestra di dialogo Sessione Utilità SPSS Production


Lavoro1 - Sessione Utilità Production
Generale
Valori di runtime
Un valore di runtime viene sostituito da un simbolo utilizzato nella sintassi. Un simbolo inizia con @ e non può contenere caratteri speciali o spazi. Se non viene specificato alcun valore predefinito, non utilizzare la parola chiave 'silent' della riga di comando durante l'esecuzione del lavoro dalla riga di comando.
Simbolo
Valore predefinito
Prompt utente
Valore virgolette
@location
\program files\samples
Where is the data file?
@datafile
demo.sav
What is the name of the data file?
@varlist
gender inccat
What variables should be used?
Esegui sessione Utilità Production
Chiudi
Aiuto

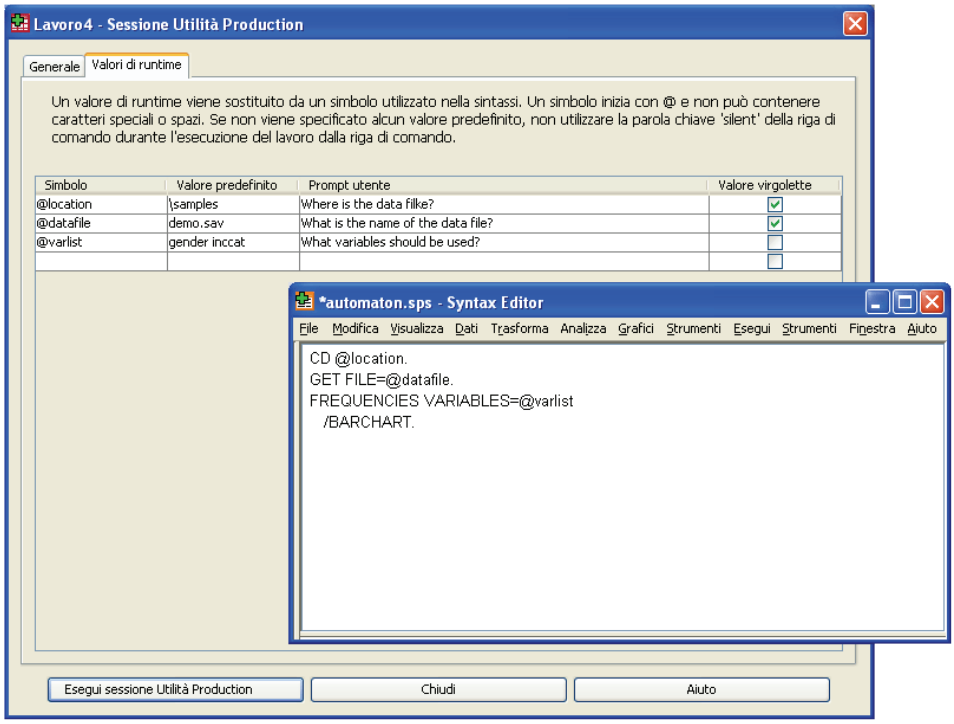
Simbolo. La stringa nel file di sintassi dei comandi che attiva la richiesta di un valore all'utente da parte della sessione Utilità SPSS Production. Il nome del simbolo deve iniziare con un carattere @ e deve essere conforme alle regole di denominazione delle variabili. Per ulteriori informazioni, vedere Nomi delle variabili in Capitolo 5 a pag. 79.

Valore default. Il valore fornito per impostazione predefinita dalla sessione dell'utilità SPSS Production se non viene specificato un valore diverso. Questo valore viene visualizzato quando nella sessione dell'utilità SPSS Production viene richiesto l'inserimento di informazioni. È possibile sostituire o modificare il valore in fase di esecuzione. Se non viene specificato alcun valore predefinito, non utilizzare la parola chiave `silent` quando si esegue la sessione Utilità SPSS Production con le opzioni della riga di comando, a meno che non si utilizzi anche l'opzione `-symbol` per specificare i valori di runtime. Per ulteriori informazioni, vedere Esecuzione di sessioni dell'utilità SPSS Production dalla riga di comando a pag. 583.

Prompt utente. L'etichetta descrittiva visualizzata quando nella sessione dell'utilità SPSS Production vengono richieste informazioni all'utente. Ad esempio, è possibile utilizzare l'espressione "Quale file dati utilizzare?" per identificare un campo che richiede l'indicazione del nome di un file dati.

Valore virgolette. Racchiude il valore predefinito o il valore immesso dall'utente tra le virgolette, ad esempio, le specificazioni dei file.

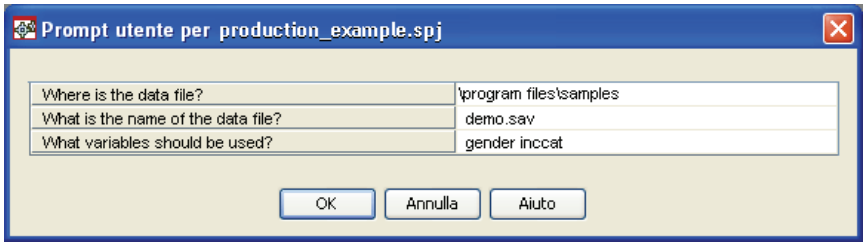
Figura 51-3
Simboli di runtime in un file della sintassi dei comandi


Lavoro4 - Sessione Utilità Production
Generale
Valori di runtime
Un valore di runtime viene sostituito da un simbolo utilizzato nella sintassi. Un simbolo inizia con @ e non può contenere caratteri speciali o spazi. Se non viene specificato alcun valore predefinito, non utilizzare la parola chiave 'silent' della riga di comando durante l'esecuzione del lavoro dalla riga di comando.
Simbolo
Valore predefinito
Prompt utente
Valore virgolette
@location
\samples
Where is the data filke?
@datafile
demo.sav
What is the name of the data file?
@varlist
gender inccat
What variables should be used?
*automaton.sps - Syntax Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Esegui Strumenti Finestra Aiuto
CD @location.
GET FILE=@datafile.
FREQUENCIES VARIABLES=@varlist
/BARCHART.
Esegui sessione Utilità Production
Chiudi
Aiuto

Prompt utente

Una sessione Utilità SPSS Production richiede di inserire valori ad ogni esecuzione della sessione contenente simboli di runtime definiti. È possibile sostituire o modificare i valori predefiniti visualizzati. Tali valori verranno sostituiti ai simboli di runtime in tutti i file di sintassi associati alla sessione Utilità SPSS Production.

Figura 51-4
Finestra di dialogo relativa alle richieste utente della sessione Utilità SPSS Production



Esecuzione di sessioni dell'utilità SPSS Production dalla riga di comando

Le opzioni della riga di comando consentono di programmare automaticamente l'esecuzione delle sessioni Utilità SPSS Production agli orari specificati grazie alle utilità di pianificazione disponibili nel sistema operativo. Il formato generico dell'argomento della riga di comando è la seguente:

```
statistics filename.spj -production
```

In base alla modalità in cui viene richiamata la sessione Utilità SPSS Production, potrebbe essere necessario includere i percorsi di directory del file eseguibile *statistics* (presente nella directory in cui è installata l'applicazione) e/o il file della sessione Utilità SPSS Production.

Per eseguire le sessioni dell'utilità SPSS Production dalla riga di comando, utilizzare le seguenti opzioni:

-production [prompt|silent]. Consente di avviare l'applicazione in modalità di produzione. Le parole chiave `prompt` e `silent` specificano se deve essere visualizzata la finestra di dialogo che richiede di immettere valori di runtime qualora vengano specificati nel lavoro. La parola chiave predefinita è `prompt` che mostra la finestra di dialogo. La parola chiave `silent` sopprime la finestra di dialogo. Utilizzando la parola chiave `silent`, è possibile definire i simboli di runtime con l'opzione `-symbol`. In caso contrario, verrà utilizzato il valore predefinito. Le opzioni `-switchserver` e `-singleseat` vengono ignorate quando si utilizza l'opzione `-production`.

-symbol <valori>. Consente di elencare le coppie simbolo-valore utilizzate nella sessione Utilità SPSS Production. Ogni nome di simbolo inizia con un carattere @. I valori che contengono spazi devono essere racchiusi tra virgolette. Le regole per includere le virgolette o gli apostrofi nelle stringhe di lettere possono variare in base ai sistemi operativi, ma racchiudere tra virgolette doppie una stringa contenente virgolette semplici o apostrofi è un'operazione che si rivela sempre efficace (ad esempio, `"'un valore racchiuso tra virgolette'"`).

Per eseguire sessioni dell'utilità SPSS Production su un server remoto nella modalità di analisi distribuita, è necessario anche specificare le informazioni di login del server:

-server <inet:nomehost:porta>. Il nome o l'indirizzo IP e il numero di porta del server. Solo per Windows.

-user <nome>. Un nome utente valido. Se è richiesto un nome di dominio, far precedere il nome utente dal nome di dominio e da una barra rovesciata (\). Solo per Windows.

-password <password>. La password dell'utente. Solo per Windows.

Esempio

```
statistics \production_jobs\prodjob1.spj -production silent -symbol @datafile /data/July_data.sav
```

- In questo esempio si presuppone che si stia eseguendo la riga di comando dalla directory di installazione; pertanto non è richiesto il percorso del file eseguibile *statistics*.
- In questo esempio si presume inoltre che la sessione Utilità SPSS Production specifica che il valore per *@datafile* deve essere racchiuso tra virgolette (casella di controllo Valore virgolette nella scheda Valori di runtime); pertanto le virgolette non sono necessarie quando si specifica il file di dati dalla riga di comando. In alternativa, è necessario specificare una stringa simile a

'/data/July_data.sav' per includere le virgolette con la specificazione dei file di dati, poiché le specificazioni devono essere inserite tra virgolette nella sintassi dei comandi.

- Il percorso della directory per la posizione della sessione Utilità SPSS Production utilizza la convenzione delle barre retroverse di Windows. In Macintosh e Linux, utilizzare le barre normali. Le barre normali nella specificazione del file di dati racchiusa tra virgolette funzioneranno su tutti i sistemi operativi quando questa stringa racchiusa tra virgolette viene inserita nel file di sintassi dei comandi e le barre normali vengono supportate nei comandi che includono le specifiche dei file, (ad esempio `GET FILE`, `GET DATA` e `SAVE`) su tutti i sistemi operativi.
- La parola chiave `silent` elimina qualsiasi richiesta dell'utente nella sessione Utilità SPSS Production e l'opzione `-symbol` inserisce il nome e la posizione del file di dati racchiuso tra virgolette ogni volta che viene visualizzato il simbolo di runtime *@datafile* nei file della sintassi dei comandi inclusi nella sessione Utilità SPSS Production.

Conversione dei file della sessione Utilità SPSS Production

I file Production Facility (.spp) creati con le versioni precedenti alla 16.0 non possono essere utilizzati con la versione 16.0 o con le versioni successive. Per i file di questo tipo di Windows e Macintosh creati con le versioni precedenti, è possibile utilizzare *prodconvert*, presente nella directory di installazione, per convertire questi file in nuovi file di sessioni Utilità SPSS Production (.spj). Eseguire *prodconvert* da una finestra dei comandi utilizzando le seguenti specifiche:

```
[installpath]\prodconvert [filepath]\filename.spp
```

dove *[installpath]* è la posizione della cartella in cui è installato SPSS Statistics e *[filepath]* è la posizione della cartella in cui si trova il file di sessioni dell'utilità SPSS Production originale. Viene creato un nuovo file con lo stesso nome ma con estensione *.spj* nella stessa cartella del file originale. (Nota: se il percorso contiene spazi, racchiudere ogni specifica di percorso e di file tra virgolette. Nei sistemi operativi Macintosh, utilizzare barre normali anziché barre rovesciate).

Limitazioni

- I formati dei grafici WMF e EMF non vengono supportati. Il formato PNG viene utilizzato al posto di questi formati.
- Le opzioni di esportazione Output (senza grafici), Solo grafici e Non pubblicare non vengono supportate. Sono inclusi tutti gli oggetti di output supportati dal formato selezionato.
- Le impostazioni relative al server remoto vengono ignorate. Per specificare le impostazioni relative al server remoto per l'analisi distribuita, è necessario eseguire la sessione Utilità SPSS Production dalla riga di comando utilizzando le relative opzioni. Per ulteriori informazioni, vedere Esecuzione di sessioni dell'utilità SPSS Production dalla riga di comando a pag. 583.
- Le impostazioni relative alla pubblicazione sul Web vengono ignorate.

Capitolo
52

Sistema di gestione dell'output

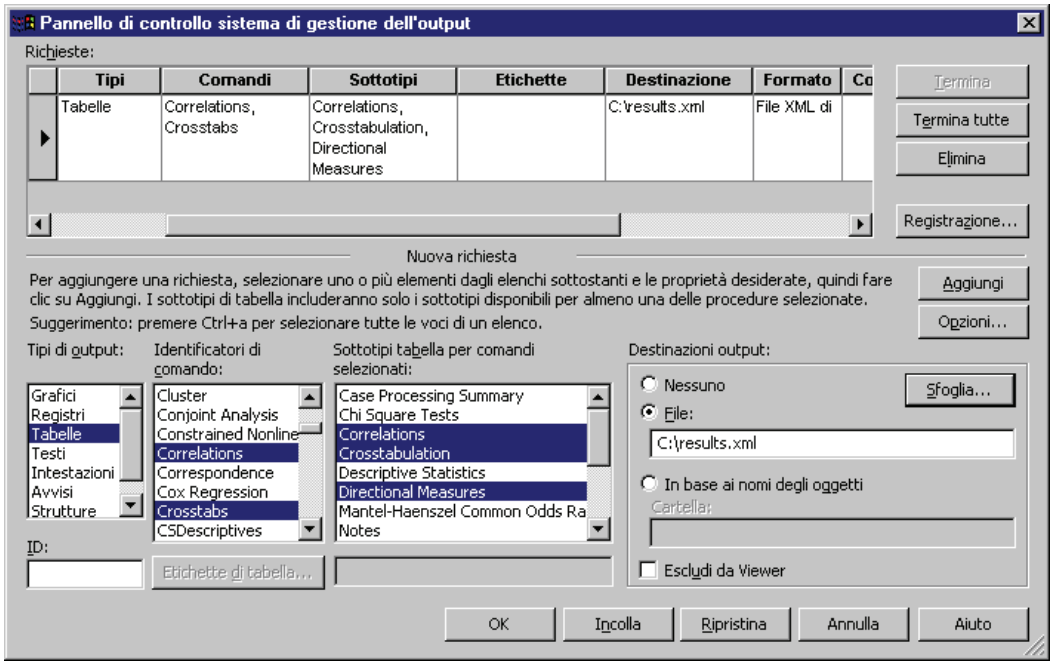
Sistema di gestione dell'output (SGO) consente di scrivere automaticamente le categorie di output selezionate in file di output in formati diversi, che includono: Word, Excel, PDF, il formato di file di dati SPSS Statistics (.sav), il formato di file del Viewer (.spv), il formato di report Web (.spw), l'XML, l'HTML e il formato di testo. Per ulteriori informazioni, vedere Opzioni SGO a pag. 591.

Per utilizzare il Pannello di controllo del sistema di gestione dell'output:

▶ Dai menu, scegliere:

Strumenti
  Pannello di controllo SGO...

Figura 52-1
Pannello di controllo sistema di gestione dell'output



È possibile usare il pannello di controllo per avviare o arrestare il trasferimento dell'output a varie destinazioni.

- Tutte le richieste SGO rimangono attive fino quando non vengono terminate o fino alla fine della sessione.
- Il file di destinazione specificato in una richiesta SGO viene reso disponibile alle altre procedure ed alle altre applicazioni solo al termine della richiesta SGO.

- Poiché i file di destinazione specificati vengono salvati nella memoria (RAM) quando la richiesta è attiva, è possibile che le richieste SGO attive che scrivono grandi volumi di output sui file esterni utilizzino molta memoria.
- Le richieste SGO sono indipendenti le une dalle altre. Lo stesso output può essere trasmesso a posizioni diverse in formati diversi, a seconda delle specifiche delle singole richieste SGO.
- L'ordine degli oggetti dell'output di una destinazione specifica varia a seconda dell'ordine e del funzionamento delle procedure che generano l'output.

Per aggiungere nuove richieste SGO

▶ Selezionare i tipi di output (tabelle, grafici, ecc.) da includere. Per ulteriori informazioni, vedere Tipi di oggetto di output a pag. 587.

▶ Selezionare i comandi da includere. Per includere tutto l'output, selezionare tutte le voci dell'elenco. Per ulteriori informazioni, vedere Identificatori di comando e sottotipi di tabelle a pag. 589.

▶ Per i comandi che producono una tabella pivot come output, selezionare i tipi di tabelle specifici da includere.

L'elenco visualizza solo le tabelle disponibili nei comandi selezionati. Quindi, l'elenco visualizza tutti ti tipi di tabelle disponibili in uno o più dei comandi selezionati. Se non sono stati selezionati comandi, vengono visualizzati tutti i tipi di tabelle. Per ulteriori informazioni, vedere Identificatori di comando e sottotipi di tabelle a pag. 589.

▶ Per selezionare le tabelle in base alle etichette del testo anziché in base ai sottotipi, fare clic su Etichette. Per ulteriori informazioni, vedere Etichette a pag. 590.

▶ Fare clic su Opzioni per specificare il formato dell'output (ad esempio file di dati di SPSS Statistics, XML, HTML). Per impostazione predefinita, viene usato il formato File XML di output. Per ulteriori informazioni, vedere Opzioni SGO a pag. 591.

▶ Specificare la destinazione di output:

- **File.** Tutto l'output selezionato viene indirizzato a un unico file.
- **Basati su nomi oggetto.** L'output viene inviato a più file di destinazione in base ai nomi degli oggetti. Viene creato un file diverso per ciascun oggetto di output, con un nome di file basato o sui nomi del sottotipo di tabella o sulle etichette delle tabelle. Immettere il nome della cartella di destinazione.
- **Nuovo file di dati.** L'output in formato di file di dati SPSS Statistics può essere inviato a un insieme di dati. Il file di dati può quindi essere usato nell'ambito della stessa sessione, ma non viene salvato a meno che non si decida esplicitamente di salvarlo al termine della sessione. Questa opzione è disponibile solo per l'output in formato di file di dati SPSS Statistics. I nomi degli insiemi di dati devono essere conformi alle regole di denominazione delle variabili. Per ulteriori informazioni, vedere Nomi delle variabili in Capitolo 5 a pag. 79.

▶ Oppure:

- Escludere l'output selezionato da Viewer. Se si seleziona Escludi da Viewer, i tipi di output della richiesta SGO non vengono visualizzati nella finestra Viewer. Se più richieste SGO attive includono gli stessi tipi di output, i tipi di output visualizzati in Viewer variano a

seconda della richiesta SGO più recente che contiene i tipi di output specificati. Per ulteriori informazioni, vedere Esclusione della visualizzazione dell'output in Viewer a pag. 597.

- Assegnare una stringa ID alla richiesta. A tutte le richieste viene automaticamente assegnato un valore ID, sebbene sia sempre possibile sostituire la stringa ID predefinita del sistema con l'ID descrittivo. Questa operazione può essere utile per identificare più facilmente più richieste attive. I valori ID assegnabili non possono iniziare con il segno Dollaro ($).

Di seguito vengono riportati una serie di consigli utili per selezionare più voci negli elenchi:

- Premere Ctrl+A per selezionare tutte le voci dell'elenco.
- Premere Maiusc ed il pulsante del mouse per selezionare più voci contigue.
- Premere Ctrl ed il pulsante del mouse per selezionare più voci non contigue.

Per terminare ed eliminare richieste SGO

Le nuove richieste SGO attive vengono visualizzate nell'elenco Richieste, a partire dalla richiesta più recente. Per modificare la larghezza delle colonne contenenti informazioni, è sufficiente selezionare e trascinare i bordi, quindi scorrere orizzontalmente l'elenco per visualizzare ulteriori informazioni sulla richiesta desiderata.

L'asterisco (*) dopo la parola Attivo nella colonna *Stato* indica che la richiesta SGO è stata creata con una sintassi di comando che comprende funzionalità non disponibili nel Pannello di comando.

Per terminare una richiesta SGO attiva specifica:

▶ Nell'elenco Richieste fare clic su qualsiasi cella della riga relativa alla richiesta.

▶ Fare clic su Termina.

Per terminare tutte le richieste SGO attive:

▶ Fare clic su Termina tutte.

Per eliminare una nuova richiesta (ossia una richiesta aggiunta ma non ancora attiva):

▶ Nell'elenco Richieste fare clic su qualsiasi cella della riga relativa alla richiesta.

▶ Fare clic su Elimina.

Nota: le richieste SGO attive vengono terminate solo quando si seleziona OK.

Tipi di oggetto di output

Esistono diversi tipi di oggetti di output:

Grafici. Sono inclusi i grafici creati con il Generatore di grafici, le procedure di rappresentazione grafica e i grafici creati dalle procedure statistiche (ad esempio, un grafico a barre creato dalla procedura Frequenze).

Intestazioni. Oggetti di testo etichettati come *titolo* nella vista riassuntiva del Viewer. Gli oggetti Intestazione non vengono inclusi nel formato di output XML.

Registri. Oggetti Testo registri. Questi oggetti contengono alcuni tipi di errori e messaggi di avviso. A seconda delle impostazioni selezionate in Opzioni (menu Modifica, Opzioni, scheda Viewer), questi oggetti possono contenere anche la sintassi dei comandi eseguita durante la sessione. Gli oggetti registro sono etichettati *Registro* nella vista riassuntiva di Viewer.

Modelli. Oggetti di output visualizzati nel Viewer modelli. Un singolo oggetto modello può contenere più viste del modello, compresi grafici e tabelle.

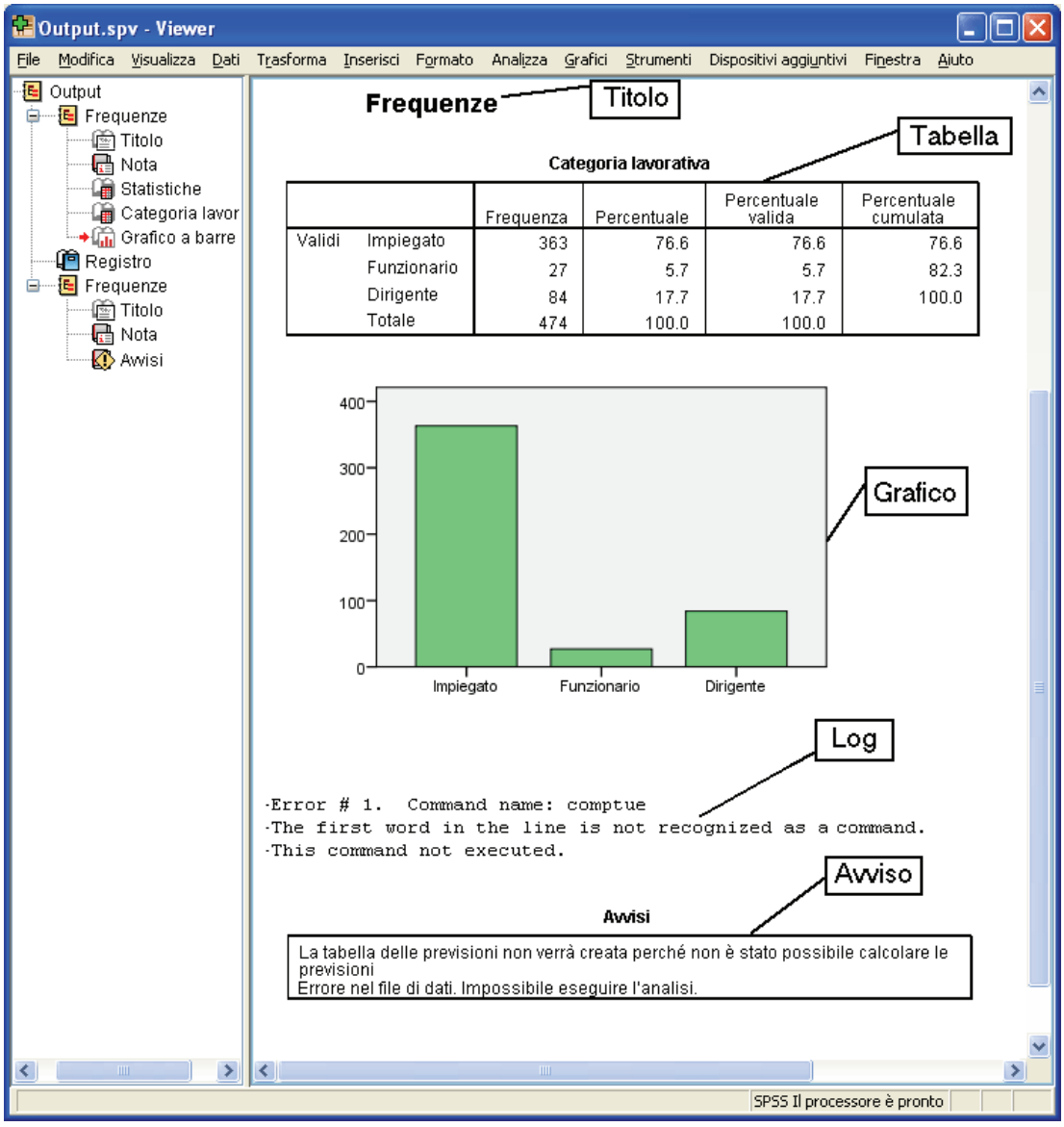
Tabelle. Le tabelle pivot in Viewer (con le tabelle Note). Le tabelle sono gli unici oggetti di output che possono essere trasmessi in formato di file di dati SPSS Statistics (.sav).

Testi. Oggetti di testo diversi da registri o titoli (compresi oggetti etichettati come *Output di testo* nella vista riassuntiva del Viewer).

Alberi. Diagrammi dei modelli di alberi creati tramite l'opzione Albero decisionale.

Avvisi. Questi oggetti contengono alcuni tipi di errori e messaggi di avviso.

Figura 52-2
Tipi di oggetti di output



Identificatori di comando e sottotipi di tabelle

Identificatori di comando

Gli identificatori di comando sono disponibili in tutte le procedure per la creazione di grafici e statistiche nonché in qualsiasi altro comando che blocchi l'output con un'intestazione identificabile nella vista riassuntiva di Viewer. Questi identificatori sono generalmente, ma non sempre, equivalenti o simili ai nomi delle procedure dei menu o dei titoli delle finestre di dialogo, che sono

generalmente, ma non sempre, simili ai nomi dei comandi associati. L'identificatore del comando per la procedura Frequenze è "Frequenze", che corrisponde anche al nome del comando associato.

Tuttavia, in alcuni casi il nome della procedura non corrisponde a quello dell'identificatore del comando e/o al nome del comando. Esempio: tutte le procedure del sottomenu Test Nonparametrici (nel menu Analizza) utilizzano lo stesso comando di base, ma l'identificatore del comando è uguale al nome del comando di base, Test non parametrici.

Sottotipi di tabella

I sottotipi di tabelle rappresentano i diversi tipi di tabelle pivot che possono essere prodotti. Alcuni sottotipi possono essere prodotti con un solo tipo di comando, mentre altri possono essere prodotti con più tipi di comandi, sebbene le tabelle abbiano lo stesso aspetto. Sebbene i nomi dei sottotipi di tabelle siano descrittivi, sono generalmente disponibili varie opzioni (soprattutto se si è selezionato un numero elevato di comandi); quindi è possibile che due sottotipi abbiano nomi simili.

Per ricercare identificatori di comandi e sottotipi di tabelle

In caso di dubbio, è possibile ricercare i nomi degli identificatori di comando e dei sottotipi delle tabelle nella finestra di Viewer.

- Eseguire la procedura per generare l'output in Viewer.
- Fare clic con il pulsante destro del mouse nella vista riassuntiva di Viewer.
- Selezionare Copia identificatore comandi SGO oppure Copia sottotipo tabelle SGO.
- Incollare l'identificatore di comando o il nome del sottotipo di tabelle copiato in un editor di testo (ad esempio una finestra dell'Editor della sintassi).

Etichette

Come alternativa ai nomi dei sottotipi di tabelle, è possibile selezionare le tabelle in base al testo visualizzato nella vista riassuntiva del Viewer. È possibile anche selezionare altri tipi di oggetto in base alle relative etichette. Le etichette consentono di distinguere tra più tabelle dello stesso tipo il cui testo riassuntivo rifletta alcuni attributi dell'oggetto di output; ad esempio i nomi o le etichette delle variabili. Il testo delle etichette può tuttavia essere influenzato da numerosi fattori:

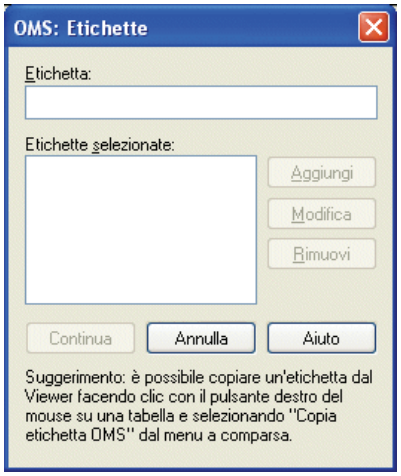
- Se è attiva la distinzione per gruppi, è possibile che all'etichetta venga aggiunta l'identificazione del gruppo di file distinto in base alla variabile categoriale.
- Le etichette che contengono informazioni sulle variabili o i valori sono influenzate dalle impostazioni delle opzioni selezionate per le etichette di output (menu Modifica, Opzioni, Etichette di output).
- Le etichette sono influenzate dall'impostazione corrente della lingua di output (menu Modifica, Opzioni, scheda Generale).

Per specificare le etichette da usare per identificare gli oggetti di output

- Nel Pannello di controllo di SGO, selezionare uno o più tipi di output, quindi uno o più tipi di comandi.

- Fare clic su Etichette.

Figura 52-3
Finestra di dialogo Etichette di SGO



- Digitare l'etichetta *esattamente* come appare nella vista riassuntiva della finestra Viewer. In alternativa, è possibile anche fare clic con il pulsante destro del mouse sulla voce nella vista riassuntiva, selezionare Copia etichetta SGO ed incollare l'etichetta copiata nel campo di testo Etichetta.
- Fare clic su Aggiungi.
- Ripetere l'operazione per ciascuna etichetta che si desidera includere.
- Fare clic su Continua.

Caratteri jolly

È possibile utilizzare un asterisco (*) per l'ultimo carattere della stringa dell'etichetta in modo che funga da carattere jolly. In questo modo vengono selezionate tutte le etichette che iniziano con la stringa specificata (ad eccezione dell'asterisco). Ciò vale solo se l'asterisco è l'ultimo carattere, poiché gli asterischi possono essere utilizzati come caratteri validi per le etichette.

Opzioni SGO

È possibile utilizzare la finestra di dialogo Opzioni SGO per:

- Specificare il formato dell'output.
- Specificare il formato immagine (per i formati di output HTML e XML di output).
- Specificare quali elementi della dimensione della tabella devono essere inclusi nella dimensione della riga.
- Per il formato di file di dati SPSS Statistics, includere una variabile che identifichi il numero della tabella sequenziale che rappresenta l'origine di ciascun caso.

Per specificare le opzioni SGO:

- Fare clic su Opzioni nel Pannello di controllo sistema di gestione dell'output.

Figura 52-4
Finestra di dialogo Opzioni SGO

Formato

Excel. Formato Excel 97-2003. Le righe, le colonne e le celle delle tabelle pivot vengono esportate come righe, colonne e celle di Excel, con tutti gli attributi di formattazione, ad esempio, con i bordi delle celle, gli stili del carattere, il colore di sfondo delle celle e così via. L'output testuale viene esportato con tutti gli attributi del carattere. Ogni riga nell'output testuale è una riga del file di Excel, con tutti i contenuti di una riga raccolti in una singola cella. I grafici, i diagrammi ad albero e le viste del modello vengono inclusi nel formato PNG.

HTML. Gli oggetti che vengono rappresentati come tabelle pivot in Viewer vengono convertiti in semplici tabelle HTML. Gli attributi TableLook (caratteristiche dei caratteri, stili dei bordi, colori, ecc.) non sono supportati. Agli oggetti di output di test viene assegnato il tag <PRE> nel file HTML. I grafici, i diagrammi ad albero e le viste dei modelli vengono esportati come file separati nel formato grafico selezionato ed incorporati per riferimento. I nomi dei file di immagine utilizzano il nome di file HTML come nome radice, seguito da un intero sequenziale che inizia con 0.

File XML di output. File XML conforme allo schema *Output SPSS*.

PDF. Tutto l'output viene esportato così come appare in Anteprima di stampa, con tutti gli attributi di formattazione invariati. Il file PDF include segnalibri che corrispondono alle voci della vista riassuntiva del Viewer.

SPSS Statistics File dati. File in formato binario. Tutti i tipi di oggetti di output diversi dalle tabelle sono esclusi. Ciascuna colonna della tabella diventa una variabile nel file di dati. Per utilizzare un file di dati creato con SGO nella stessa sessione, è necessario terminare la richiesta SGO attiva prima di aprire il file di dati. Per ulteriori informazioni, vedere Trasmissione dell'output ai file di dati in formato SPSS Statistics a pag. 597.

Testo. Testo separato da spazi. L'output viene scritto come testo, con l'output della tabella allineato agli spazi per i caratteri con passo fisso. I grafici, i diagrammi ad albero e le viste modello sono esclusi.

Testo con tabulazioni. Testo delimitato da tabulazioni. Le tabulazioni delimitano gli elementi delle colonne delle tabelle per l'output che viene visualizzato come tabelle pivot in Viewer. Le righe dei blocchi di testo vengono scritte come tali senza che vengano inserite tabulazioni per dividerle nei punti più opportuni. I grafici, i diagrammi ad albero e le viste modello sono esclusi.

File Viewer. Si tratta dello stesso formato utilizzato quando si salva il contenuto di una finestra del Viewer.

File di Reporto Web. Questo formato del file di output è destinato a essere utilizzato con Predictive Enterprise Services. Si tratta essenzialmente dello stesso formato del Viewer di SPSS Statistics, con la differenza che i diagrammi ad albero vengono salvati come immagini statiche.

Word/RTF. Le tabelle pivot vengono esportate come tabelle Word con tutti gli attributi di formattazione, ad esempio bordi delle celle, stili del carattere e colori dello sfondo. L'output testuale viene esportato come HTML formattato. I grafici, i diagrammi ad albero e le viste del modello vengono inclusi nel formato PNG.

Immagini grafiche

Nel formato HTML e nel formato XML di output, è possibile includere grafici, diagrammi ad albero e viste modelli come file di immagine. Per ciascun grafico e/o albero viene creato un file di immagine separato.

- Nel formato HTML, i tag standard `<IMG SRC='filename'>` vengono inclusi nel documento HTML per ciascun file di immagine.
- Per il formato XML di output, il file XML contiene un elemento `grafico` con un attributo `ImageFile` del formato generico `<grafico imageFile="filepath/filename"/>` per ciascun file di immagine.
- I file d'immagine vengono salvati in una sottodirectory a parte (cartella). Il nome della sottodirectory corrisponde al nome del file di destinazione senza estensioni, al quale viene aggiunto *_files*. Ad esempio, se il file di destinazione è *julydata.htm*, la sottodirectory delle immagini verrà chiamata *julydata_files*.

Formato. I file di immagine disponibili sono PNG, JPG, EMF, BMP e VML.

- il formato EMF (Metafile avanzato) è disponibile solo nei sistemi operativi Windows.
- Il formato di immagine VML è disponibile solo per il formato HTML.
- Il formato di immagine VML nono crea file di immagine separati. Il codice VML che consente di visualizzare l'immagine è incorporato nel codice HTML.
- Il formato di immagine VML non include diagrammi ad albero.

Dimensioni. È possibile scalare le dimensioni dell'immagine nell'intervallo 10% - 200%.

Includi imagemap. Per il formato HTML, questa opzione crea strumenti imagemap che mostrano informazioni relative ad alcuni elementi del grafico, ad esempio il valore del punto selezionato in un grafico lineare o della barra selezionata in un grafico a barre.

Tabelle pivot

Per l'output in formato tabella pivot, è possibile specificare gli elementi delle dimensioni che devono essere visualizzati nelle colonne. Tutte le altre dimensioni vengono visualizzate sulle righe. Nel formato di file di dati SPSS Statistics, le colonne delle tabelle diventano variabili, mentre le righe diventano casi.

- Se si specificano elementi con dimensioni diverse per le colonne, questi vengono annidati nell'ordine in cui sono presenti nell'elenco. Per il formato di file di dati SPSS Statistics, i nomi delle variabili vengono creati utilizzano gli elementi delle colonne annidate. Per ulteriori informazioni, vedere Nomi di variabili nei file dati generati da SGO a pag. 604.
- Se la tabella non contiene elementi relativi alle dimensioni, questi vengono visualizzati sulle righe.
- I pivot tabella specificati in questo contesto non hanno alcun effetto sulle tabelle visualizzate in Viewer.

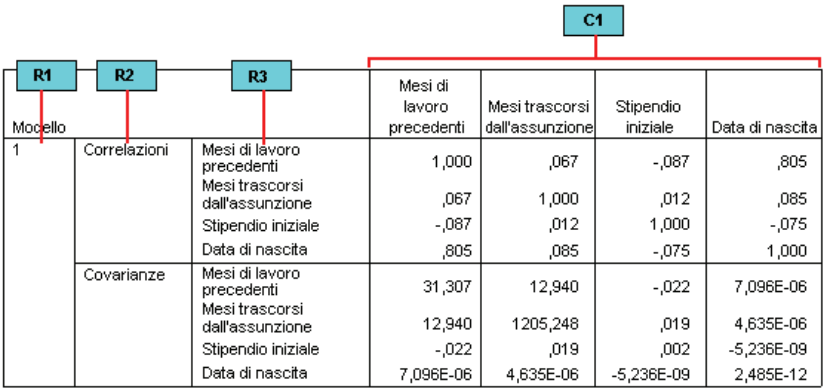
Ciascuna dimensione di una tabella - riga, colonna, strato - possono contenere più elementi o nessun elemento. Ad esempio una tavola di contingenza contiene una dimensione in un'unica riga e una dimensione in una sola colonna; ciascuna di queste dimensioni contiene una delle variabili usate nella tabella. È possibile usare un argomento relativo alla posizione o "nomi" di dimensioni per specificare le dimensioni che si desidera inserire nella dimensione della colonna.

Tutte le dimensioni nelle righe. Crea una singola riga per ciascuna tabella. Per i file di dati in formato SPSS Statistics, ciò significa che ciascuna tabella rappresenta un caso singolo e che tutti gli elementi della tabella sono variabili.

Elenco posizioni. Un argomento relativo alla posizione è generalmente rappresentato da una lettera che indica la posizione predefinita dell'elemento - C per la colonna, R per la riga o L per lo strato - seguita da un numero intero positivo che indica la posizione predefinita della dimensione. R1, ad esempio, indica la dimensione della riga più esterna.

- Per specificare più dimensioni, è necessario separare ciascuna dimensione con uno spazio—ad esempio R1 C2.
- Se la lettera della dimensione è seguita da ALL, significa che tutti gli elementi della dimensione sono disposti nell'ordine predefinito. CALL, ad esempio, è uguale al comportamento predefinito poiché utilizza tutti gli elementi della colonna nell'ordine predefinito per creare le colonne.
- CALL RALL LALL (o RALL CALL LALL, etc.) inserisce tutte le dimensioni nelle colonne. Nel formato di file di dati SPSS Statistics, ciò crea una riga/un caso per tabella nel file dati.

Figura 52-5
Argomenti relativi alla posizione di righe e colonne



Modello			Mesi di lavoro precedenti	Mesi trascorsi dall'assunzione	Stipendio iniziale	Data di nascita
1	Correlazioni	Mesi di lavoro precedenti	1,000	,067	-,087	,805
		Mesi trascorsi dall'assunzione	,067	1,000	,012	,085
		Stipendio iniziale	-,087	,012	1,000	-,075
		Data di nascita	,805	,085	-,075	1,000
	Covarianze	Mesi di lavoro precedenti	31,307	12,940	-,022	7,096E-06
		Mesi trascorsi dall'assunzione	12,940	1205,248	,019	4,635E-06
		Stipendio iniziale	-,022	,019	,002	-5,236E-09
		Data di nascita	7,096E-06	4,635E-06	-5,236E-09	2,485E-12

Elenco nomi di dimensione. In alternativa agli argomenti relativi alla posizione, è possibile utilizzare i "nomi" delle dimensioni rappresentati dalle etichette di testo visualizzate nella tabella. Ad un esempio una tavola di convergenza semplice bidimensionale contiene una dimensione di una singola riga e una dimensione di una singola colonna, entrambe con etichette basate sulle variabili contenute nelle dimensioni specificate, oltre ad una dimensione di uno strato singolo etichettato *Statistiche* (se è stato impostato l'italiano come lingua di output).

- I nomi delle dimensioni possono variare a seconda della lingua di output e/o delle impostazioni che influiscono sulla visualizzazione dei nomi delle variabili e/o delle etichette nelle tabelle.
- Ciascun nome di dimensione deve essere racchiuso tra virgolette semplici o doppie. Per specificare nomi di più dimensioni, è necessario aggiungere uno spazio tra ciascun nome.

Le etichette associate alle dimensioni non sono sempre ovvie.

Per visualizzare tutti gli elementi dimensione e le relative etichette di una tabella pivot

▶ Attivare (con un doppio clic) del mouse la tabella in Viewer.

▶ Dai menu, scegliere:

Visualizza
  Mostra tutto

e/o

▶ Se Pivoting non è visualizzato, dai menu scegliere:

Pivot
  Pivoting

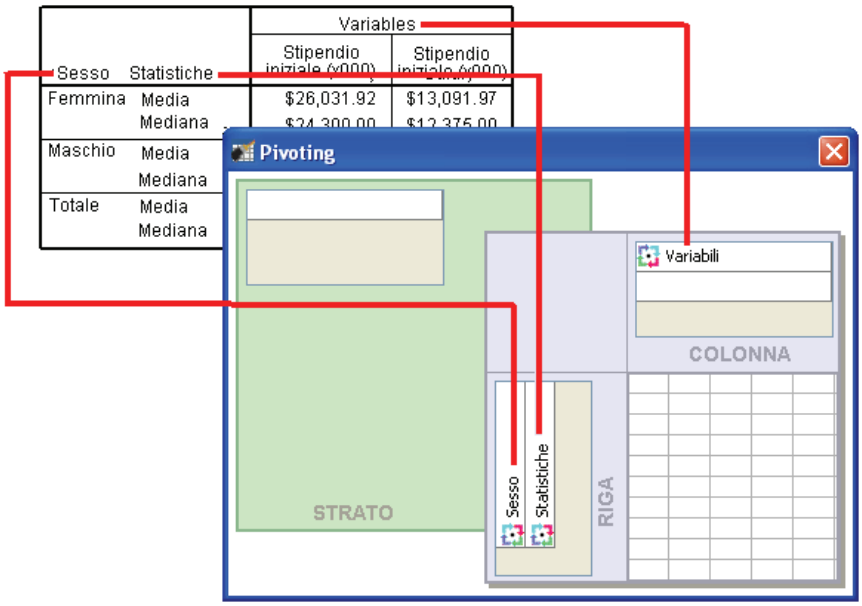
Le etichette degli elementi vengono visualizzate sui pivoting.

Figura 52-6
Nomi delle dimensioni visualizzate nella tabella e nei pivot

Tabella senza visualizzazione delle etichette delle dimensioni

		Stipendio iniziale (x000)	Stipendio iniziale (x000)
Femmina	Media	$26,031.92	$13,091.97
	Mediana	$24,300.00	$12,375.00
Maschio	Media	$41,441.78	$20,301.40
	Mediana	$32,850.00	$15,750.00
Totale	Media	$34,419.57	$17,016.09
	Mediana	$28,875.00	$15,000.00

Tabella con visualizzazione delle etichette delle dimensioni



Registrazione

È possibile registrare un'attività SGO in un registro in formato XML o testo.

- Il registro tiene traccia di tutte le nuove richieste SGO della sessione, ma non include le richieste SGO che erano già attive prima della richiesta di un registro.
- Il file di registro corrente termina se si specifica un nuovo file di registro o se si deseleziona l'opzione Registra attività SGO.

Per specificare una registrazione SGO:

▶ Fare clic su Registrazione nel Pannello di controllo sistema di gestione dell'output.

Esclusione della visualizzazione dell'output in Viewer

La casella di controllo Escludi da Viewer ha effetto su tutto l'output selezionato nella richiesta SGO, poiché evita che tale output venga visualizzato nella finestra del Viewer. Questo processo è spesso utile per le sessioni Utilità SPSS Production che generano molto output e per i quali non si desidera visualizzare i risultati in formato di documento Viewer (file .spv). L'opzione può anche essere usata per nascondere oggetti di output specifici che non si desiderano visualizzare senza dover trasmettere l'output ad un file e formato esterno.

Per nascondere oggetti di output specifici senza trasmettere altri output ad un file esterno:

- Creare una richiesta SGO che identifichi l'output indesiderato.
- Selezionare Escludi da Viewer.
- Per la destinazione di output selezionare File, lasciando il campo File vuoto.
- Fare clic su Aggiungi.

L'output selezionato viene escluso da Viewer, mentre tutti gli altri output vengono visualizzati in Viewer nel formato standard.

Trasmissione dell'output ai file di dati in formato SPSS Statistics

I file di dati in formato SPSS Statistics sono costituiti dalle variabili delle colonne e dai casi delle righe, che rappresenta essenzialmente la modalità con cui le tabelle pivot vengono convertite in file di dati:

- Le colonne della tabella sono le variabili del file dati. I nomi delle variabili validi vengono creati utilizzando le etichette delle colonne.
- Le etichette delle righe della tabella diventano i nomi di variabili generiche (*Var1*, *Var2*, *Var3*, ecc.) nel file di dati. Il valore di queste variabili è rappresentato dalle etichette delle righe nella tabella.
- Nel file dati vengono automaticamente incluse tre variabili identificatore della tabella: *Command_*, *Subtype_* e *Label_*. Tutte e tre sono variabili stringa. Le prime due variabili corrispondono al comando e agli identificatori del sottotipo. Per ulteriori informazioni, vedere Identificatori di comando e sottotipi di tabelle a pag. 589. *Label_* contiene il testo del titolo della tabella.
- Le righe della tabella diventano i casi del file di dati.

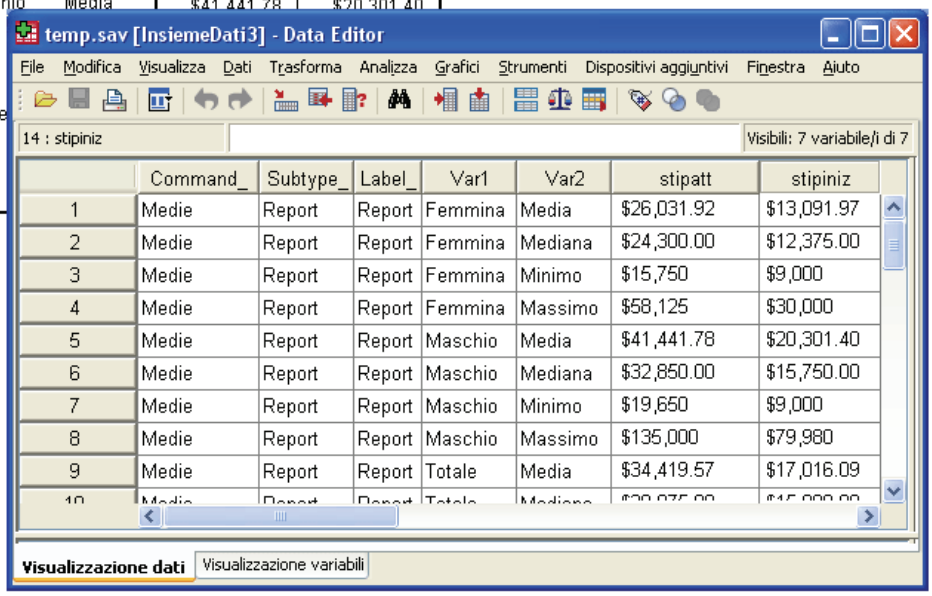
Esempio: tabella bidimensionale singola

Nel caso più semplice, rappresentato da una tabella bidimensionale singola, le colonne della tabella diventano variabili, mentre le righe diventano i casi del file di dati.

Figura 52-7
Tabella bidimensionale singola

Report

Sesso		Stipendio attuale (x000)	Stipendio iniziale (x000)
Femmina	Media	$26.031,92	$13.091,97
	Mediana	$24.300,00	$12.375,00
	Minimo	$15.750,00	$9.000,00
	Massimo	$58.125,00	$30.000,00
Maschio	Media	$41.441,78	$20.301,40
Totale			


temp.sav [InsiemeDati3] - Data Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntivi Finestra Aiuto
14 : stipiniz
Visibili: 7 variabile/i di 7
Visualizzazione dati
Visualizzazione variabili

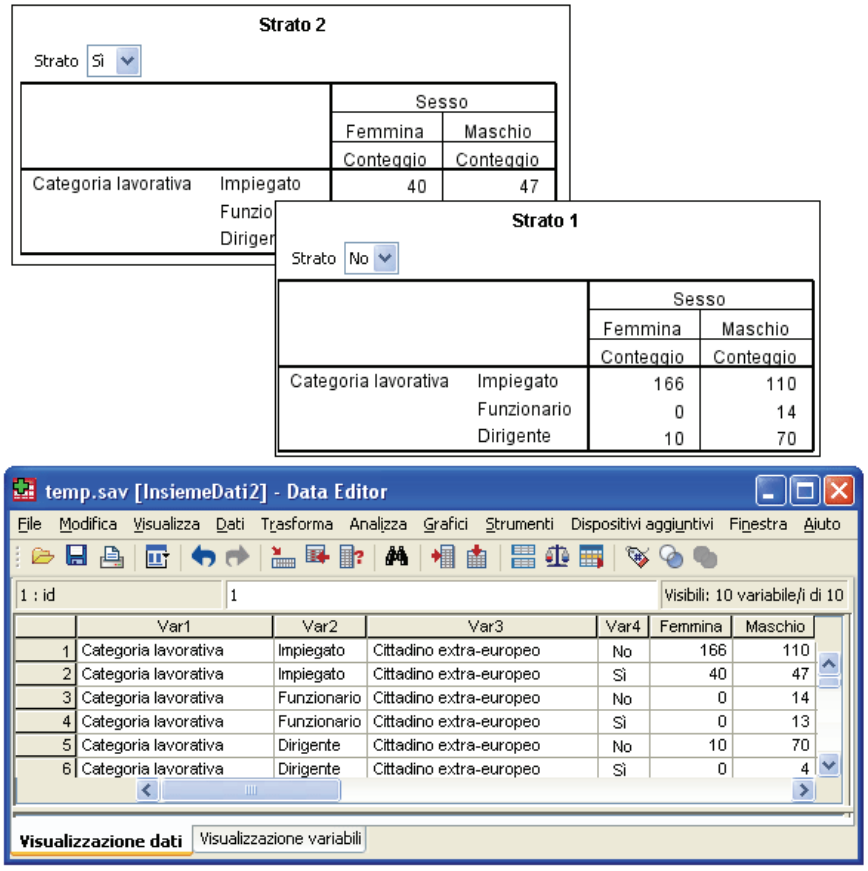
	Command_	Subtype_	Label_	Var1	Var2	stipatt	stipiniz
1	Medie	Report	Report	Femmina	Media	$26,031.92	$13,091.97
2	Medie	Report	Report	Femmina	Mediana	$24,300.00	$12,375.00
3	Medie	Report	Report	Femmina	Minimo	$15,750	$9,000
4	Medie	Report	Report	Femmina	Massimo	$58,125	$30,000
5	Medie	Report	Report	Maschio	Media	$41,441.78	$20,301.40
6	Medie	Report	Report	Maschio	Mediana	$32,850.00	$15,750.00
7	Medie	Report	Report	Maschio	Minimo	$19,650	$9,000
8	Medie	Report	Report	Maschio	Massimo	$135,000	$79,980
9	Medie	Report	Report	Totale	Media	$34,419.57	$17,016.09
10	Medie	Report	Report	Totale	Mediana		

- Le prime tre variabili identificano la tabella di origine in base al sottotipo di comando e all'etichetta.
- Ai due elementi che definivano le righe nella tabella, ossia i valori della variabile *Sesso* e le misure statistiche, vengono assegnati nomi di variabili generici *Var1* e *Var2*. Questi variabili sono entrambe variabili stringa.
- Le etichette delle colonne della tabella vengono usate per creare nomi di variabili valide. In questo caso, i nomi delle variabili sono basati sulle etichette delle variabili delle tre variabili scala riassunte nella tabella. Se le variabili non hanno un'etichetta variabile definita o si sceglie di visualizzare i nomi delle variabili anziché le etichette delle variabili come etichette delle colonne della tabella, i nomi delle variabili nel nuovo file saranno gli stessi del file dati di origine.

Esempio: Tabelle con strati

Oltre a righe e colonne una tabella può contenere anche un terzo elemento: lo strato.

Figura 52-8
Tabella con strati



Strato 2

Strato: Sì

		Sesso	
		Femmina	Maschio
		Conteggio	Conteggio
Categoria lavorativa	Impiegato	40	47
	Funzio		
	Diriger		

Strato 1

Strato: No

		Sesso	
		Femmina	Maschio
		Conteggio	Conteggio
Categoria lavorativa	Impiegato	166	110
	Funzionario	0	14
	Dirigente	10	70

temp.sav [InsiemeDati2] - Data Editor

File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntivi Finestra Aiuto

1 : id | 1 | Visibili: 10 variabile/i di 10

	Var1	Var2	Var3	Var4	Femmina	Maschio
1	Categoria lavorativa	Impiegato	Cittadino extra-europeo	No	166	110
2	Categoria lavorativa	Impiegato	Cittadino extra-europeo	Sì	40	47
3	Categoria lavorativa	Funzionario	Cittadino extra-europeo	No	0	14
4	Categoria lavorativa	Funzionario	Cittadino extra-europeo	Sì	0	13
5	Categoria lavorativa	Dirigente	Cittadino extra-europeo	No	10	70
6	Categoria lavorativa	Dirigente	Cittadino extra-europeo	Sì	0	4

Visualizzazione dati | Visualizzazione variabili

- La variabile denominata *Classificazione per minoranza* definisce gli strati nella tabella. Nel file di dati vengono create altre due variabili: la prima identifica lo strato, mentre la seconda identifica le categorie degli elementi dello strato.
- Al pari delle variabili create con le righe, le variabili create con gli strati sono variabili stringa con nomi generici (formato dal prefisso *Var* seguito da un numero progressivo).

File dati creati con tabelle multiple

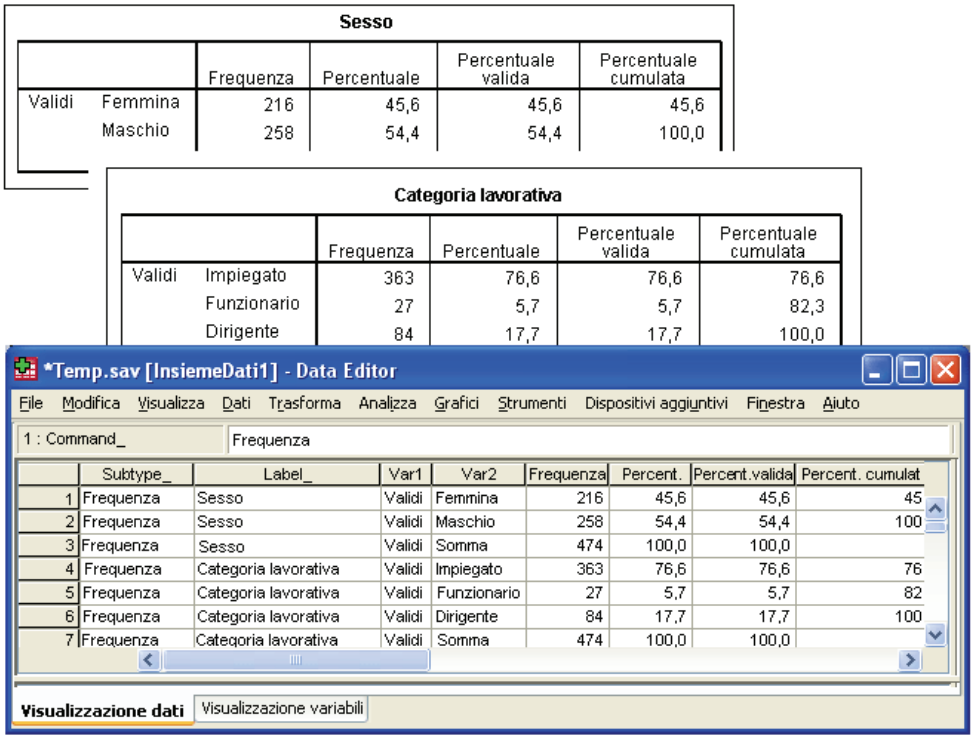
Se si associano più tabelle allo stesso file dati, ciascuna tabella viene aggiunta al file dati con modalità simili a quelle utilizzate per unire i file di dati aggiungendo casi da un file di dati ad un altro (menu Dati, Unisci file, Aggiungi casi).

- Ciascuna tabella successiva aggiungerà sempre i casi al file di dati.
- Se le etichette delle colonne della tabella differiscono, è possibile anche che ciascuna tabella aggiunga variabili al file di dati utilizzando anche i valori mancanti dei casi di altre tabelle che non hanno etichette simili per le colonne.

Esempio: tabelle multiple con le stesse etichette colonne

Le tabelle multiple che contengono le stesse etichette per le colonne forniscono generalmente i file di dati più utili (ossia file di dati che non richiedono ulteriori manipolazioni). Due o più tabelle della frequenza generate tramite la procedura Frequenza possono avere, ad esempio, etichette identiche per le colonne.

Figura 52-9
Due tabelle con etichette identiche per le colonne



Sesso

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	Femmina	216	45,6	45,6	45,6
	Maschio	258	54,4	54,4	100,0

Categoria lavorativa

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	Impiegato	363	76,6	76,6	76,6
	Funzionario	27	5,7	5,7	82,3
	Dirigente	84	17,7	17,7	100,0

*Temp.sav [InsiemeDati1] - Data Editor

File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntivi Finestra Aiuto

1 : Command_ | Frequenza

	Subtype_	Label_	Var1	Var2	Frequenza	Percent.	Percent.valida	Percent. cumulat
1	Frequenza	Sesso	Validi	Femmina	216	45,6	45,6	45
2	Frequenza	Sesso	Validi	Maschio	258	54,4	54,4	100
3	Frequenza	Sesso	Validi	Somma	474	100,0	100,0	
4	Frequenza	Categoria lavorativa	Validi	Impiegato	363	76,6	76,6	76
5	Frequenza	Categoria lavorativa	Validi	Funzionario	27	5,7	5,7	82
6	Frequenza	Categoria lavorativa	Validi	Dirigente	84	17,7	17,7	100
7	Frequenza	Categoria lavorativa	Validi	Somma	474	100,0	100,0	

Visualizzazione dati | Visualizzazione variabili

- La seconda tabella fornisce ulteriori casi (righe) al file di dati ma nessuna variabile nuova perché le etichette delle colonne sono identiche, quindi grosse sezioni di dati mancanti.
- Sebbene i valori di *Command_* e *Subtype_* siano identici, il valore di *Label_* identifica la tabella di origine per ciascun gruppo di casi perché le due tabelle relative alla frequenza hanno titoli diversi.

Esempio: tabelle multiple con etichette di colonna diverse

Nel file di dati viene creata una nuova variabile per ciascuna etichetta di colonna univoca delle tabelle associate al file di dati. Questo processo genera blocchi di valori mancanti se le tabelle contengono etichette colonne diverse.

Figura 52-10
Due tabelle con etichette di colonna diverse

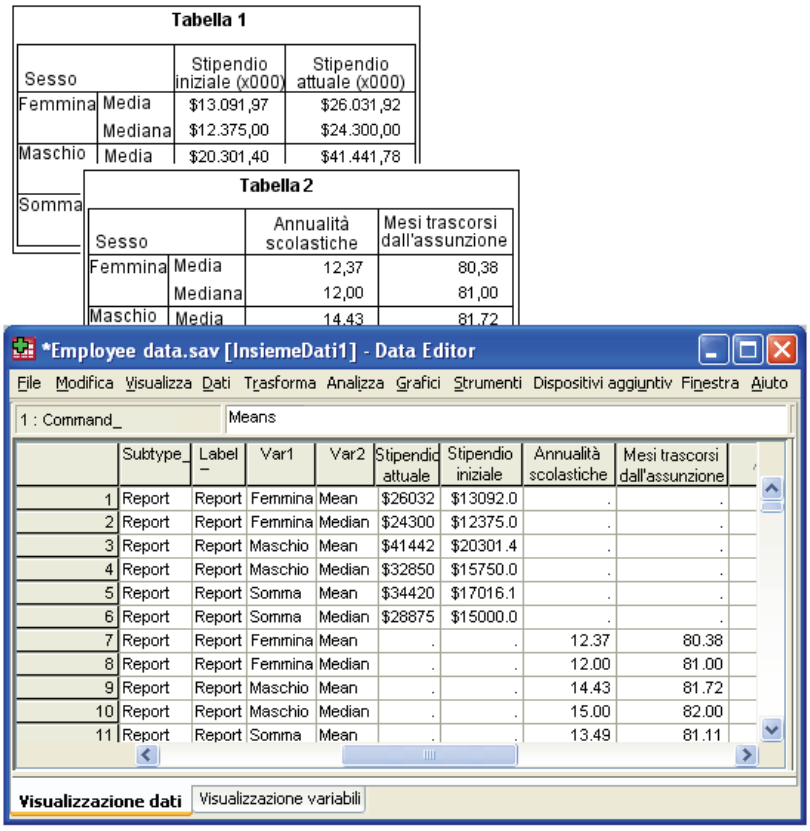


Tabella 1

Sesso		Stipendio iniziale (x000)	Stipendio attuale (x000)
Femmina	Media	$13.091,97	$26.031,92
	Mediana	$12.375,00	$24.300,00
Maschio	Media	$20.301,40	$41.441,78
Somma			

Tabella 2

Sesso		Annualità scolastiche	Mesi trascorsi dall'assunzione
Femmina	Media	12,37	80,38
	Mediana	12,00	81,00
Maschio	Media	14,43	81,72

*Employee data.sav [InsiemeDati1] - Data Editor
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntiv Finestra Aiuto
1 : Command_
Means
Visualizzazione dati
Visualizzazione variabili

	Subtype_	Label_	Var1	Var2	Stipendio attuale	Stipendio iniziale	Annualità scolastiche	Mesi trascorsi dall'assunzione
1	Report	Report	Femmina	Mean	$26032	$13092.0	.	.
2	Report	Report	Femmina	Median	$24300	$12375.0	.	.
3	Report	Report	Maschio	Mean	$41442	$20301.4	.	.
4	Report	Report	Maschio	Median	$32850	$15750.0	.	.
5	Report	Report	Somma	Mean	$34420	$17016.1	.	.
6	Report	Report	Somma	Median	$28875	$15000.0	.	.
7	Report	Report	Femmina	Mean	.	.	12.37	80.38
8	Report	Report	Femmina	Median	.	.	12.00	81.00
9	Report	Report	Maschio	Mean	.	.	14.43	81.72
10	Report	Report	Maschio	Median	.	.	15.00	82.00
11	Report	Report	Somma	Mean	.	.	13.49	81.11

- La prima tabella ha due colonne denominate rispettivamente *Stipendio iniziale* e *Stipendio attuale* che non sono presenti nella seconda tabella e che generano valori mancanti per le variabili dei casi della seconda tabella.
- La seconda tabella ha due colonne denominate rispettivamente *Livello di educazione* e *Mesi di servizio* che non sono presenti nella prima tabella e che provocano valori mancanti per le variabili dei casi della prima tabella.
- Differenze nella variabili simile a quelle dell'esempio possono verificarsi anche per tabelle dello stesso sottotipo. Le tabelle dell'esempio sono infatti entrambe dello stesso sottotipo.

Esempio: file di dati non creati con tabelle multiple

I file di dati non vengono creati per le tabelle che non hanno lo stesso numero di elementi riga. Il numero di righe non deve essere necessariamente lo stesso, ma il numero di *elementi* riga che si trasformano nel file di dati deve sempre essere uguale. Esempio: una tabella incrociata con due variabili e una tabella incrociata con tre variabili contengono numeri di elementi riga diversi, poiché la variabile "strato" è nidificata nella variabile riga della vista della tabella predefinita con tre variabili.

Figura 52-11
Tabelle con numero di elementi riga diversi

Categoria lavorativa * Tabella di contingenza Genere

		Sesso		
		Femmina	Maschio	Totale
Categoria lavorativa	Impiegato	206	157	363
	Funzionario	0	27	27
	Dirigente	10	74	84
Totale		216	258	474

Categoria lavorativa * Genere * Tabella di contingenza Cittadini extra-europei

			Sesso		
Cittadini extra-europei			Femmina	Maschio	Totale
No	Categoria lavorativa	Impiegato	166	110	276
		Funzionario	0	14	14
		Dirigente	10	70	80
	Totale		176	194	370
Sì	Categoria lavorativa	Impiegato	40	47	87
		Funzionario	0	13	13
		Dirigente	0	4	4
	Totale		40	64	104

Uso degli elementi colonna per il controllo delle variabili nei file di dati

Nella finestra di dialogo Opzioni del Pannello di controllo SGO è possibile specificare gli elementi dimensione per le colonne, che verranno successivamente usati per creare le variabili nei file di dati generati. Questo processo è analogo ad effettuare il pivoting in Viewer.

Esempio: la procedura Frequenze produce una tabella statistica descrittiva in cui i dati statistici sono contenuti nelle righe; la procedure Descrittive invece produce una tabella statistica descritta in cui i dati statistici sono contenuti nelle colonne. Per poter includere entrambe le tabelle in un file di dati in modo significativo, è necessario modificare la dimensione delle colonne di uno dei tipi di tabella.

Poiché entrambe le tabelle usano il nome “Statistiche” per la dimensione statistiche, è possibile inserire le statistiche della tabella della procedura Frequenze nella colonne selezionando semplicemente “Statistiche” (tra virgolette) nell’elenco dei nomi delle dimensioni nella finestra di dialogo Opzioni di SGO.

Figura 52-12
Finestra di dialogo Opzioni SGO

Figura 52-13
Inserimento di tipi di tabelle diversi in un file di dati tramite gli elementi della dimensione pivot

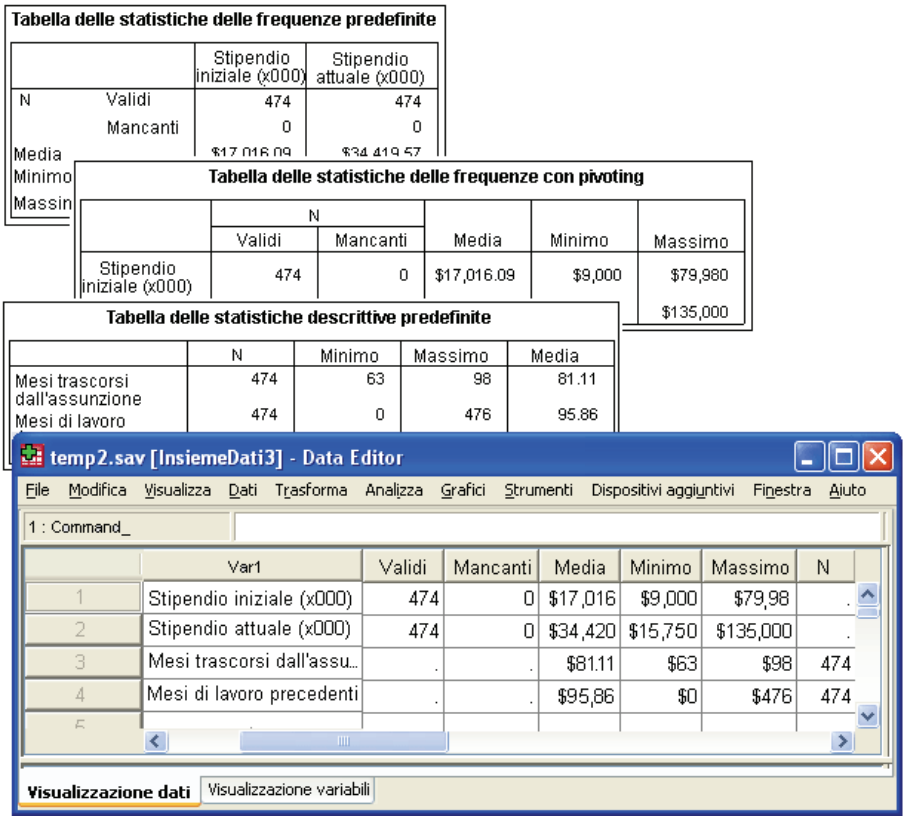


Tabella delle statistiche delle frequenze predefinite

		Stipendio iniziale (x000)	Stipendio attuale (x000)
N	Validi	474	474
	Mancanti	0	0
Media		$17,016.09	$34,419.57
Minimo		$9,000	$15,750
Massimo		$79,980	$135,000

Tabella delle statistiche delle frequenze con pivoting

	N		Media	Minimo	Massimo
	Validi	Mancanti			
Stipendio iniziale (x000)	474	0	$17,016.09	$9,000	$79,980
					$135,000

Tabella delle statistiche descrittive predefinite

	N	Minimo	Massimo	Media
Mesi trascorsi dall'assunzione	474	63	98	81.11
Mesi di lavoro	474	0	476	95.86

temp2.sav [InsiemeDati3] - Data Editor

File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Dispositivi aggiuntivi Finestra Aiuto

1 : Command_

	Var1	Validi	Mancanti	Media	Minimo	Massimo	N
1	Stipendio iniziale (x000)	474	0	$17,016	$9,000	$79,98	.
2	Stipendio attuale (x000)	474	0	$34,420	$15,750	$135,000	.
3	Mesi trascorsi dall'assu...	.	.	$81.11	$63	$98	474
4	Mesi di lavoro precedenti	.	.	$95,86	$0	$476	474
5							

Visualizzazione dati | Visualizzazione variabili

Alcune delle variabili avranno dati mancanti poiché le strutture delle tabelle non sono esattamente le stesse dei dati statistici contenuti nelle colonne.

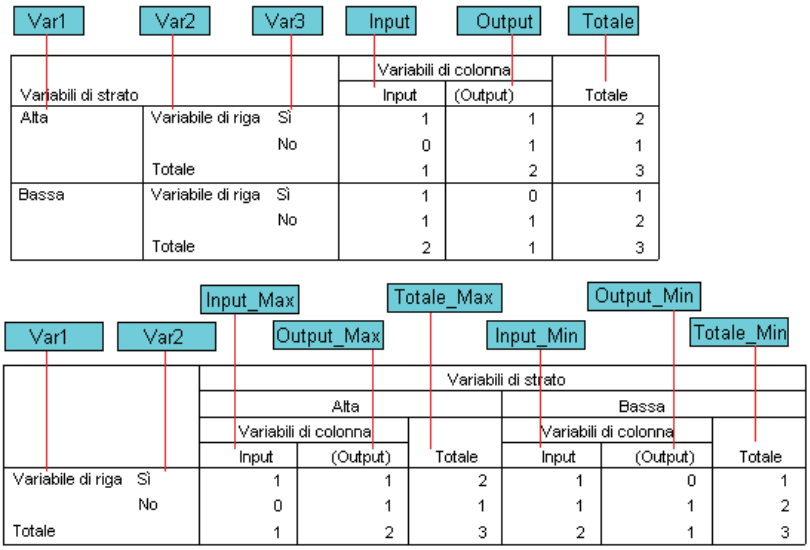
Nomi di variabili nei file dati generati da SGO

SGO genera nomi di variabili validi ed univoci sulla base delle etichette delle colonne:

- Agli elementi riga e strato vengono assegnati nomi di variabili generiche —ovvero il prefisso *Var* seguito da un numero progressivo.
- I caratteri non consentiti per i nomi delle variabili (come spazi e parentesi, ecc.) vengono eliminati. Esempio: “Questa etichetta (colonna)” diventa una variabile denominata *QuestaEtichettaColonna*.
- Se l’etichetta inizia con un carattere ammesso per i nomi delle variabili ma non utilizzabile come primo carattere (ad esempio un numero), “@” viene inserito come prefisso. Esempio: “2°” diventa una variabile denominata *@2°*.

- I trattini di sottolineatura o i punti alla fine delle etichette vengono cancellati dai nomi delle variabili risultanti. I trattini di sottolineatura alla fine delle variabili generate automaticamente *Command_*, *Subtype_* e *Label_* non vengono rimossi.
- Se la dimensione colonna contiene più elementi, i nomi delle variabili vengono creati combinando le etichette delle categorie con i trattini di sottolineato tra le etichette delle categorie stesse. Le etichette dei gruppi non vengono incluse. Esempio. se *VarB* è nidificata sotto a *VarA* nelle colonne, le variabili risultanti sono *CatA1_CatB1*, non *VarA_CatA1_VarB_CatB1*.

Figura 52-14
Nomi di variabili creati dagli elementi della tabella



Struttura della tabella OXML

Output XML (OXML) è un XML conforme allo schema *spss-output*. Per informazioni dettagliate sullo schema, vedere la sezione Schema dell'output nella Guida.

- Il comando SGO e gli identificatori dei sottotipi vengono usati come valori del comando e attributi subType in OXML. Esempio:

```
<command text=Frequencies command=Frequencies...>
 <pivotTable text=Gender label=Gender subType=Frequencies...>
```

- I valori degli attributi command e subType non sono influenzati dalla lingua dell'output o dalle impostazioni di visualizzazione impostate per i nomi/le etichette delle variabili o le etichette dei valori/del valore.
- XML distingue tra lettere minuscole e maiuscole. Il valore dell'attributo subType di "frequenze" *non* è uguale al valore dell'attributo subType di "Frequenze."

- Tutte le informazioni visualizzate nella tabella sono contenute nei valori degli attributi in OXML. A livello di cella singola, OXML è costituito da due elementi "vuoti" che contengono gli attributi ma nessun "contenuto" oltre a quello dei valori degli attributi.
- La struttura della tabella in OXML è rappresentata riga per riga; gli elementi che rappresentano le colonne sono nidificati nelle righe mentre le singole celle sono nidificate negli elementi delle colonne:

```
<pivotTable...>
 <dimension axis='row'...>
  <dimension axis='column'...>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
  </dimension>
 </dimension>
  ...
 </pivotTable>
```

L'esempio precedente è una rappresentazione semplificata della struttura che mostra le relazioni discendente/antenato di questi elementi, ma non mostra necessariamente una relazione padre/figlio a causa dei livelli di elementi nidificati.

La figura che segue mostra una tabella semplice della frequenza e la rappresentazione d'output completa della tabella in XML.

Figura 52-15
Tabella della frequenza semplice

Sesso

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	Femmina	216	45,6	45,6	45,6
	Maschio	258	54,4	54,4	100,0
	Totale	474	100,0	100,0	

Figura 52-16
Output XML della tabella della frequenza semplice

```
<?xml version=1.0 encoding=UTF-8?>
<outputTreeoutputTree xmlns=http://xml.spss.com/spss/oms
 xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
 xsi:schemaLocation=http://xml.spss.com/spss/oms
 http://xml.spss.com/spss/oms/spss-output-1.0.xsd>
 <command text=Frequencies command=Frequencies
 displayTableValues=label displayOutlineValues=label
 displayTableVariables=label displayOutlineVariables=label>
 <pivotTable text=Gender label=Gender subType=Frequencies
  varName=gender variable=true>
  <dimension axis=row text=Gender label=Gender
```

```
varName=gender variable=true>
<group text=Valid>
 <group hide=true text=Dummy>
  <category text=Female label=Female string=f
   varName=gender>
   <dimension axis=column text=Statistics>
    <category text=Frequency>
     <cell text=216 number=216/>
    </category>
    <category text=Percent>
     <cell text=45.6 number=45.569620253165 decimals=1/>
    </category>
    <category text=Valid Percent>
     <cell text=45.6 number=45.569620253165 decimals=1/>
    </category>
    <category text=Cumulative Percent>
     <cell text=45.6 number=45.569620253165 decimals=1/>
    </category>
   </dimension>
  </category>
  <category text=Male label=Male string=m varName=gender>
   <dimension axis=column text=Statistics>
    <category text=Frequency>
     <cell text=258 number=258/>
    </category>
    <category text=Percent>
     <cell text=54.4 number=54.430379746835 decimals=1/>
    </category>
    <category text=Valid Percent>
     <cell text=54.4 number=54.430379746835 decimals=1/>
    </category>
    <category text=Cumulative Percent>
     <cell text=100.0 number=100 decimals=1/>
    </category>
   </dimension>
  </category>
 </group>
 <category text=Total>
  <dimension axis=column text=Statistics>
   <category text=Frequency>
    <cell text=474 number=474/>
   </category>
   <category text=Percent>
    <cell text=100.0 number=100 decimals=1/>
   </category>
   <category text=Valid Percent>
    <cell text=100.0 number=100 decimals=1/>
   </category>
  </dimension>
 </category>
</group>
</dimension>
```

```
 </pivotTable>
 </command>
</outputTree>
```

L'esempio mostra che anche una piccola tabella semplice può produrre una grande quantità di codice XML. Ciò è dovuto in parte al fatto che il codice XML contiene informazioni non immediatamente visibili nella tabella originale, ossia informazioni che possono non essere disponibili nella tabella originale oltre ad una certa quantità di informazioni ridondanti.

- I contenuti della tabella che vengono visualizzati nella tabella pivot di Viewer sono contenuti negli attributi di testo. Esempio:

```
<command text=Frequencies command=Frequencies...>
```

- Gli attributi di testo sono influenzati sia dalla lingua dell'output che dalle impostazioni di visualizzazione dei nomi/delle etichette delle variabili e delle etichette dei valori/del valore. Nell'esempio, il valore dell'attributo testo varia a seconda della lingua di output,mentre il valore dell'attributo del comando rimane sempre lo stesso indipendentemente dalla lingua di output.
- Mentre le variabili o i valori vengono usati per le etichette di righe o colonne, in XML contengono un attributo testo ed uno o più valori di attributi aggiuntivi. Esempio:

```
<dimension axis=row text=Gender label=Gender varName=gender>
  ...<category text=Female label=Female string=f varName=gender>
```

- Nel caso di una variabile numerica, ci sarebbe un attributo numero anziché stringa. L'attributo etichetta è presente solo se per la variabile o i valori sono stati definite etichette.
- Gli elementi <cell> contengono i valori delle celle per i numeri che contengono l'attributotesto e uno o più valori attributo. Esempio:

```
<cell text=45.6 number=45.569620253165 decimals=1/>
```

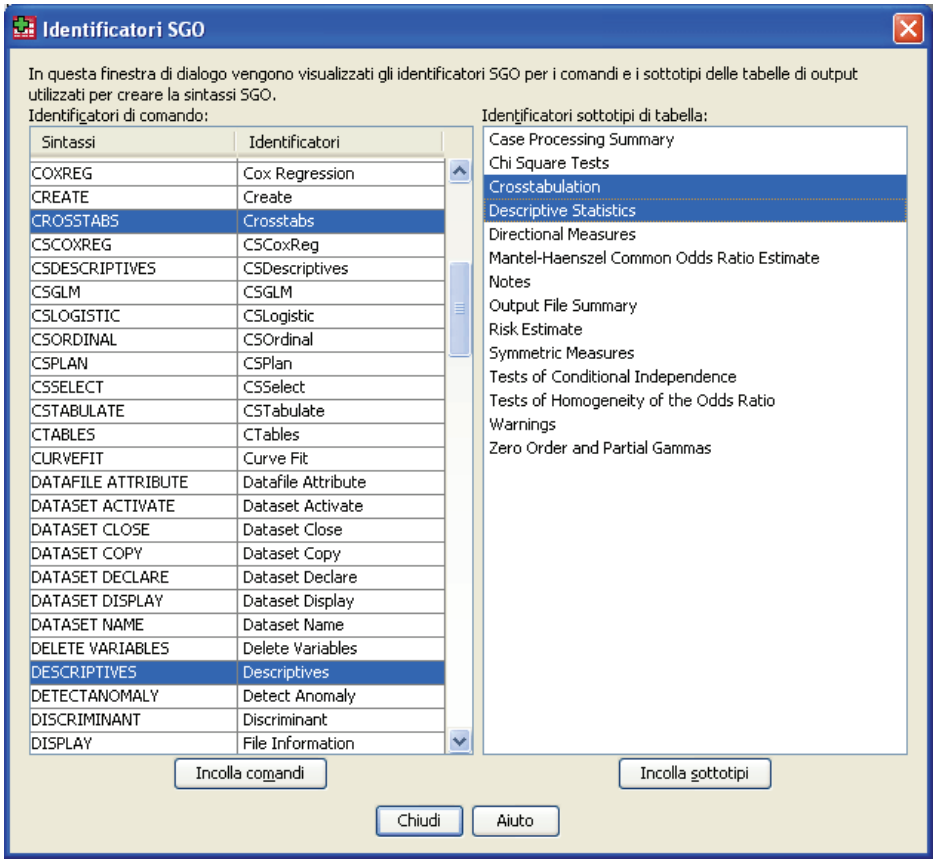
L'attributo numero rappresenta il valore numero effettivo non arrotondato, mentre l'attributo decimali indica il numero di posizioni decimali visualizzato nella tabella.

- Poiché le colonne sono nidificate nelle righe, l'elemento categoria che identifica ciascuna colonna viene ripetuto per tutte le righe. Ad esempio, poiché le statistiche sono visualizzate nelle colonne, l'elemento <category text=Frequency> appare tre volte in XML: una volta per la riga Maschio, una volta per la riga Femmina e una volta per la riga dei totali.

Identificatori SGO

La finestra di dialogo Identificatori SGO semplifica l'utilizzo della sintassi dei comandi di SGO poiché consente di incollare il comando selezionato e gli identificatori del sottotipo in una finestra della sintassi dei comandi.

Figura 52-17
Finestra di dialogo Identificatori SGO



Per utilizzare la finestra di dialogo Identificatori SGO

- Dai menu, scegliere:

  Strumenti
  Identificatori SGO...

- Selezionare uno o più degli identificatori di comandi o di sottotipi. Per selezionare più identificatori in ogni elenco, fare clic su di essi tenendo premuto Ctrl.

- Fare clic su Incolla comandi e/o su Incolla sottotipi.
  - L'elenco dei sottotipi disponibili varia in base ai comandi correntemente selezionati. Se sono selezionati più comandi, l'elenco dei sottotipi è rappresentato dall'unione di tutti i sottotipi disponibili per ognuno dei comandi selezionati. Se non sono selezionati comandi, l'elenco includerà tutti i sottotipi.
  - Gli identificatori vengono incollati nella finestra della sintassi dei comandi designata in corrispondenza della posizione del cursore corrente. Se non sono aperte finestre della sintassi dei comandi, ne verrà aperta automaticamente una nuova.

- Ogni identificatore di comando e/o sottotipo viene racchiuso tra virgolette quando viene incollato, poiché la sintassi del comando SGO le richiede.
- Gli elenchi di identificatori per le parole chiave COMMANDS e SUBTYPES devono essere racchiusi tra parentesi quadre, come è illustrato nell'esempio seguente:

```
/IF COMMANDS=['Crosstabs' 'Descriptives']
    SUBTYPES=['Crosstabulation' 'Descriptive Statistics']
```

Copia degli identificatori SGO dalla vista riassuntiva del Viewer

È possibile copiare e incollare gli identificatori del comando SGO e dei sottotipi dalla vista riassuntiva del Viewer.

▶ Nella vista riassuntiva fare clic con il pulsante destro del mouse su una voce.

▶ Selezionare Copia identificatore comandi SGO oppure Copia sottotipo tabelle SGO.

Questo metodo si differenzia dall'utilizzo della finestra Identificatori SGP per il fatto che l'identificatore copiato non viene incollato automaticamente in una finestra della sintassi dei comandi, ma viene semplicemente copiato negli Appunti, dai quali è possibile incollarlo in qualsiasi posizione. Dato che i valori degli identificatori di comandi e sottotipi sono identici ai comandi e ai valori degli attributi di sottotipo corrispondenti nell'output in formato XML (OXML), questo metodo di copia e incolla può risultare utile quando si scrivono trasformazioni XSLT.

Copia di etichette SGO

Anziché gli identificatori, è possibile copiare le etichette e utilizzarle con la parola chiave LABELS. Le etichette consentono di differenziare più grafici o più tabelle dello stesso tipo nei quali il testo dell'output rappresenta un attributo dell'oggetto di output specifico, ad esempio i nomi o le etichette delle variabili. Il testo delle etichette può tuttavia essere influenzato da numerosi fattori:

- Se è attiva la distinzione per gruppi, è possibile che all'etichetta venga aggiunta l'identificazione del gruppo di file distinto in base alla variabile categoriale.
- Le etichette che includono informazioni sulle variabili o i valori sono influenzate dalle impostazioni di visualizzazione per i nomi e le etichette delle variabili e per i valori e le etichette di valori nella vista riassuntiva (menu Modifica, Opzioni, scheda Etichette di output).
- Le etichette sono influenzate dall'impostazione corrente della lingua di output (menu Modifica, Opzioni, scheda Generale).

Per copiare le etichette SGO:

▶ Nella vista riassuntiva fare clic con il pulsante destro del mouse su una voce.

▶ Selezionare Copia etichetta SGO.

Allo stesso modo degli identificatori di comandi e di sottotipi, le etichette devono essere racchiuse tra virgolette e l'elenco completo deve essere racchiuso tra parentesi quadre, come è illustrato nell'esempio seguente:

```
/IF LABELS=['Employment Category' 'Education Level']
```

Appendice

A

Convertitore di sintassi dei comandi TABLES e IGRAPH

Se è disponibile un file della sintassi dei comandi che contiene la sintassi TABLES e si desidera convertirlo nella sintassi CTABLES, e/o se si desidera convertire IGRAPH in GGRAPH, è possibile usare una semplice utilità appositamente progettata per semplificare il processo di conversione. Tuttavia, esistono differenze di funzionalità significative fra TABLES e CTABLES e fra IGRAPH e GGRAPH. Probabilmente, l'utilità non sarà in grado di convertire alcuni dei processi di sintassi TABLES e IGRAPH, oppure la sintassi CTABLES e GGRAPH generata produrrà tabelle e grafici diversi dagli originali prodotti mediante i comandi TABLES e IGRAPH. Nella maggior parte dei casi è possibile modificare la sintassi convertita in modo da ottenere tabelle che corrispondono all'originale.

Questa utilità è progettata per:

- Creare un nuovo file di sintassi da un file esistente. Il file di sintassi originale non viene modificato.
- Convertire solo i comandi TABLES e IGRAPH contenuti nel file di sintassi. Gli altri comandi nel file non vengono modificati.
- Mantenere la sintassi TABLES e IGRAPH originale con i commenti.
- Identificare l'inizio e la fine di ciascun blocco di conversione con dei commenti.
- Identificare i comandi della sintassi TABLES e IGRAPH che non è stato possibile convertire.
- Convertire i file di sintassi dei comandi che rispettano le regole di sintassi delle modalità interattiva e produzione.

L'utilità non è in grado di convertire i comandi che contengono degli errori. Inoltre, valgono le limitazioni indicate di seguito.

Limitazioni del comando TABLES

In alcuni casi è possibile che l'utilità converta i comandi TABLES in modo errato, compresi i comandi TABLES che contengono:

- I nomi di variabili tra parentesi che iniziano con le lettere "sta" o "lab" nel sottocomando TABLES se la variabile è racchiusa tra parentesi, ad esempio `var1 by (statvar) by (labvar)`. Questi vengono interpretati come parole chiave (STATISTICS) e (LABELS).
- Sottocomandi SORT che usano le abbreviazioni A o D per indicare un ordine crescente o decrescente. Questi verranno interpretati come nomi di variabili.

L'utilità non è in grado di convertire i comandi TABLES che contengono:

- Errori di sintassi.
- Sottocomandi OBSERVATION che si riferiscono a un intervallo di variabili che utilizza la parola chiave TO (ad esempio var01 TO var05).
- Stringhe di lettere spezzate in segmenti separati dal segno più (ad esempio TITLE "My" + "Title").
- Chiamate a macro che, in assenza di un'espansione macro, rappresentano una sintassi TABLES non valida. Poiché il convertitore non espande le chiamate alle macro, le considera come parte della sintassi TABLES standard.

L'utilità non converte i comandi TABLES contenuti nelle macro. Il processo di conversione non ha alcun effetto sulle macro.

Limitazioni del comando IGRAPH

IGRAPH è stato notevolmente modificato nella versione 16. In seguito a tali modifiche, alcuni sottocomandi e parole chiave nella sintassi IGRAPH creata prima di questa versione potrebbero non essere rispettati. Vedere la sezione IGRAPH in *Command Syntax Reference* per la cronologia completa delle revisioni.

L'utilità di conversione può generare della sintassi addizionale che viene archiviata nella parola chiave INLINETEMPLATE all'interno della sintassi GGRAPH. Questa parola chiave viene creata solo dalprogramma di conversione. La sua sintassi non è modificabile dall'utente.

Uso dell'utilità di conversione

Il programma di conversione, *SyntaxConverter.exe*, è reperibile nella directory di installazione. Il programma può essere eseguito da un prompt dei comandi. La forma generica di un comando è:

```
syntaxconverter.exe [path]/inputfilename.sps [path]/outputfilename.sps
```

Questo comando deve essere eseguito dalla directory di installazione.

Se il nome di una delle directory contiene degli spazi, è necessario racchiudere l'intero percorso e nome file tra virgolette come mostra il seguente esempio:

```
syntaxconverter.exe /myfiles/oldfile.sps /new files/newfile.sps
```

Regole della sintassi dei comandi in modalità interattiva e Production

Il programma di conversione è in grado di convertire file di comando che usano regole di sintassi delle modalità interattive o Production.

Interattive. Le regole della sintassi della modalità interattiva sono:

- Ciascun comando inizia su una nuova riga.
- Ciascun comando deve terminare con un punto (.).

Modalità Production. L'utilità Production Facility e i comandi accessibili tramite il comando `INCLUDE` su un file di comando diverso usano le regole della sintassi della modalità Production:

- Ogni comando deve iniziare nella prima colonna di una nuova riga.
- Le righe di continuazione devono essere rientrate di almeno uno spazio.
- Il punto alla fine del comando è facoltativo.

Se i file dei comandi usano le regole della sintassi della modalità Production e non contengono punti alla fine di ciascun comando, è necessario includere lo switch per la riga di comando `-b` (o `/b`) per eseguire *SyntaxConverter.exe*, ad esempio:

```
syntaxconverter.exe -b /myfiles/oldfile.sps /myfiles/newfile.sps
```

Script SyntaxConverter (solo Windows)

In Windows è possibile anche eseguire il convertitore di sintassi con lo script *SyntaxConverter.wwd*, situato nella directory *Samples* della directory di installazione.

▶ Dai menu, scegliere:
Strumenti
  Esegui script...

▶ Passare alla directory *Samples* e selezionare *SyntaxConverter.wwd*.

Viene aperta una finestra di dialogo semplice in cui è possibile specificare i nomi e le posizioni dei file delle sintassi dei comandi vecchi e nuovi.

Indice

Access (Microsoft), 18
affidabilità di divisione a metà
  in analisi di affidabilità, 487, 489
affidabilità di Spearman-Brown
  in analisi di affidabilità, 489
aggiunta di etichette di gruppo, 243
aggregazione dei dati, 188
  funzioni di aggregazione, 191
  nomi di variabili ed etichette, 192
aiuto in linea, 11
  Statistics Coach, 10
alfa di Cronbach
  in analisi di affidabilità, 487, 489
algoritmi, 11
allineamento, 86, 218, 524
  nell'Editor dei dati, 86
  output, 218, 524
allocazione della memoria
  in analisi Cluster TwoStep, 431
analisi a risposta multipla
  Risposte multiple: Frequenze, 468
  Risposte multiple: Tavole di contingenza, 470
  tabelle di frequenza, 468
  tavole di contingenza, 470
analisi cluster
  Cluster con metodo delle k-medie, 442
  Cluster gerarchica, 436
  efficienza, 444
Analisi cluster TwoStep, 428
  grafici, 433
  opzioni, 431
  salvataggio in un file di lavoro, 434
  salvataggio in un file esterno, 434
  statistiche, 434
analisi dei dati, 9
  passaggi fondamentali, 9
Analisi del vicino più vicino, 389
  opzioni, 400
  output, 399
  partizioni, 396
  salvataggio di variabili, 398
  selezione delle funzioni, 395
  vicini, 393
  vista del modello, 401
analisi della varianza
  in ANOVA univariata, 333
  in medie, 317
  in regressione lineare, 369
  in stima di curve, 380
analisi delle componenti principali, 419, 422
Analisi di affidabilità, 487
  coefficiente di correlazione intraclasse, 489
  correlazione e covarianze inter-item, 489
  descrittive, 489
  esempio, 487
  funzioni aggiuntive del comando, 491
  Kuder-Richardson 20, 489
  statistiche, 487, 489
  Tabella ANOVA, 489
  Test di additività di Tukey, 489
  $T^{quadrato}$ di Hotelling, 489
Analisi discriminante, 411
  coefficienti di funzione, 414
  criteri, 415
  definizione di intervalli, 413
  Distanza di Mahalanobis, 415
  esempio, 411
  esportazione di informazioni dei modelli, 417
  funzioni aggiuntive del comando, 418
  grafici, 416
  Lambda di Wilks, 415
  matrice di covarianza, 416
  matrici, 414
  metodi discriminanti, 415
  metodi stepwise, 411
  opzioni di visualizzazione, 415–416
  probabilità a priori, 416
  salvataggio di variabili di classificazione, 417
  selezione di casi, 413
  statistiche, 411, 414
  statistiche descrittive, 414
  V di Rao , 415
  valori mancanti, 416
  variabili di raggruppamento, 411
  variabili indipendenti, 411
Analisi fattoriale, 419
  cenni generali, 419
  convergenza, 422–423
  descrittive, 421
  esempio, 419
  formato di visualizzazione dei coefficienti, 425
  funzioni aggiuntive del comando, 426
  grafici dei pesi fattoriali, 423
  metodi di estrazione, 422
  metodi di rotazione, 423
  punteggi fattoriali, 424
  selezione di casi, 420
  statistiche, 419, 421
  valori mancanti, 425
analisi serie storiche
  previsione, 382
  previsione di casi, 382
anni, 525
  valori a due cifre, 525

annotazioni, 251, 259–261
  grafici, 512
  rinumerazione, 261
  simboli, 251, 260
ANOVA
  in ANOVA univariata, 333
  in GLM univariato, 339
  in medie, 317
  modello, 341
ANOVA univariata, 333
  confronti multipli, 335
  contrasti, 334
  contrasti polinomiali, 334
  funzioni aggiuntive del comando, 338
  opzioni, 337
  statistiche, 337
  test Post Hoc, 335
  valori mancanti, 337
  variabili fattore, 333
anticipo, 142, 172
apertura di file, 14–18, 32
  controllo dei percorsi file predefiniti, 535
  file delimitati da tabulazioni, 14
  file di dati, 14–15
  file di dati testuali, 32
  file di dBASE, 14, 17
  file di Excel, 14, 16
  file di fogli elettronici, 14, 16
  File di Lotus 1-2-3, 14
  File Stata, 17
  file SYSTAT, 14
asimmetria
  in cubi OLAP, 321
  in descrittive, 295
  in esplora, 299
  in frequenze, 290
  in medie, 317
  in Report: Riepiloghi per colonne, 482
  in Report: Riepiloghi per righe, 477
  in Riassumi, 312
Assegnazione multipla
  opzioni, 541
assegnazioni
  ricerca nell'Editor dei dati, 97
associazione lineare
  in tavole di contingenza, 305
attributi
  attributi variabili personalizzati, 87
attributi di variabili, 86–87
  copiare e incollare, 86–87
  personalizzati, 87
attributi personalizzati, 87
attributi variabili personalizzati, 87
autovalori
  in analisi fattoriale, 421–422
  in regressione lineare, 369
barra di stato, 6
barre degli strumenti, 544, 546–547
  creazione, 544, 546
  creazione di nuovi strumenti, 547
  mostrare e nascondere, 544
  personalizzazione, 544, 546
  visualizzazione in finestre diverse, 546
Bonferroni
  in ANOVA univariata, 335
  in GLM, 345
bontà di adattamento
  in Regressione ordinale, 376
bordi, 254, 261
  visualizzazione dei bordi nascosti, 261

C di Dunnett
  in ANOVA univariata, 335
  in GLM, 345
calcolo di variabili, 135
  calcolo di nuove variabili stringa, 137
campionamento
  campione casuale, 196
campione casuale, 24
  database, 24
  selezione, 196
  seme dei numeri casuali, 139
campione di controllo
  nell'analisi del vicino più vicino, 396
campione di training
  nell'analisi del vicino più vicino, 396
campioni dipendenti, 459, 464
caratteri, 100, 220, 257
  nella vista riassuntiva, 220
  nell'Editor dei dati, 100
casi, 95, 199
  inserimento di nuovi casi, 95
  ordinamento, 180
  peso, 197
  ricerca di duplicati, 124
  ricerca nell'Editor dei dati, 97
  ristrutturazione in variabili, 199
  selezione di sottoinsiemi, 193, 195–196
casi duplicati (record)
  ricerca e funzioni di filtro, 124
casi filtrati, 99
  nell'Editor dei dati, 99
categoria di riferimento
  in GLM, 343
categorizzazione, 127
celle nelle tabelle pivot, 252, 260–261
  formati, 252
  larghezze, 261
  mostrare, 247
  nascondere, 247
centratura dell'output, 218, 524
chi-quadrato, 447
  associazione lineare, 305

correzione di continuità di Yates, 305
in tavole di contingenza, 305
intervallo atteso., 449
opzioni, 449
Pearson, 305
per l'indipendenza, 305
rapporto di verosimiglianza, 305
statistiche, 449
test esatto di Fisher, 305
test per un campione, 447
valori attesi, 449
valori mancanti, 449
chi-quadrato del rapporto di verosimiglianza
in Regressione ordinale, 376
in tavole di contingenza, 305
Chi-quadrato di Pearson
in Regressione ordinale, 376
in tavole di contingenza, 305
classifica di casi, 151
percentili, 152
punteggi esponenziali, 152
ranghi frazionari, 152
valori a pari merito, 153
classificazione
nella Curva ROC, 502
Cluster con metodo delle k-medie
cenni generali, 442
cluster di appartenenza, 445
criteri di convergenza, 444
distanze tra cluster, 445
efficienza, 444
esempi, 442
funzioni aggiuntive del comando, 446
iterazioni, 444
metodi, 442
salvataggio di informazioni del cluster, 445
statistiche, 442, 445
valori mancanti, 445
Cluster gerarchica, 436
cluster di appartenenza, 439–440
dendrogrammi, 440
esempio, 436
funzioni aggiuntive del comando, 441
grafici a stalattite, 440
matrici delle distanze, 439
metodi di raggruppamento, 438
misure di distanza, 438
misure di similarità, 438
orientamento del grafico, 440
programmi di agglomerazione, 439
raggruppamento dei casi in cluster, 436
salvataggio di nuove variabili, 440
statistiche, 436, 439
trasformazione di misure, 438
trasformazione di valori, 438
variabili, 436
codifica locale, 281
coefficiente alfa
in analisi di affidabilità, 487, 489
coefficiente di contingenza
in tavole di contingenza, 305
coefficiente di correlazione dei ranghi
in correlazioni bivariate, 351
coefficiente di correlazione di Spearman
in correlazioni bivariate, 351
in tavole di contingenza, 305
coefficiente di correlazione intraclasse
in analisi di affidabilità, 489
coefficiente di correlazione *r*
in correlazioni bivariate, 351
in tavole di contingenza, 305
coefficiente di dispersione (COD)
in Statistiche di rapporto, 500
coefficiente di incertezza
in tavole di contingenza, 305
coefficiente di rischio
in tavole di contingenza, 305
coefficiente di variazione (COV)
in Statistiche di rapporto, 500
coefficienti beta
in regressione lineare, 369
coefficienti di regressione.
in regressione lineare, 369
collegamento
in Regressione ordinale, 374
colonna totale
nei report, 483
colonne, 261–262
modifica della larghezza nelle tabelle pivot, 261
selezione nelle tabelle pivot, 262
colore di sfondo, 257
colori nelle tabelle pivot, 254
bordi, 254
comandi di estensione
finestre di dialogo personalizzate, 572
compressione delle categorie, 127
condizioni, 137
confronti multipli
in ANOVA univariata, 335
confronti multipli post hoc, 335
confronto di gruppi
in cubi OLAP, 324
confronto di variabili
in cubi OLAP, 324
conteggio atteso
in tavole di contingenza, 308
conteggio delle occorrenze, 140
conteggio osservato
in tavole di contingenza, 308
contrasti
in ANOVA univariata, 334
in GLM, 343
contrasti di deviazione
in GLM, 343

contrasti di Helmert
in GLM, 343
contrasti differenza
in GLM, 343
contrasti polinomiali
in ANOVA univariata, 334
in GLM, 343
contrasti ripetuti
in GLM, 343
contrasti semplici
in GLM, 343
controllo pagina
nei report di riepilogo per colonne, 484
nei report di riepilogo per righe, 478
convergenza
in analisi fattoriale, 422–423
nel cluster con metodo delle K-medie, 444
convertitore di sintassi, 612
coppie di variabili, 199
creazione, 199
Correlazione di Pearson
in correlazioni bivariate, 351
correlazione di Pearson
in tavole di contingenza, 305
correlazioni
di ordine zero, 355
in correlazioni bivariate, 351
in Correlazioni parziali, 354
in tavole di contingenza, 305
Correlazioni bivariate
coefficienti di correlazione, 351
funzioni aggiuntive del comando, 353
livello di significatività, 351
opzioni, 353
statistiche, 353
valori mancanti, 353
correlazioni di ordine zero
in Correlazioni parziali, 355
Correlazioni parziali, 354
correlazioni di ordine zero, 355
funzioni aggiuntive del comando, 356
in regressione lineare, 369
opzioni, 355
statistiche, 355
valori mancanti, 355
correzione di continuità di Yates
in tavole di contingenza, 305
costruzione di termini, 341, 378
creazione di grafici, 505
cronologia iterazioni
in Regressione ordinale, 376
CTABLES
conversione della sintassi del comando TABLES in CTABLES, 612
Cubi OLAP, 320
statistiche, 321
titoli, 325
curtosi
in cubi OLAP, 321
in descrittive, 295
in esplora, 299
in frequenze, 290
in medie, 317
in Report: Riepiloghi per colonne, 482
in Report: Riepiloghi per righe, 477
in Riassumi, 312
Curva ROC, 502
statistiche e grafici, 504

d
in tavole di contingenza, 305
d di Somers
in tavole di contingenza, 305
DATA LIST, 65
confronto con il comando GET DATA, 65
database, 18, 21, 23–24, 26, 29, 31
aggiornamento, 52
aggiunta di nuovi campi a una tabella, 59
aggiunta di nuovi record (casi) a una tabella, 60
campionamento casuale, 24
conversione di variabili stringa in variabili numeriche, 29
creazione di nuova tabella, 62
creazione di relazioni, 23
definizione di criteri, 24
definizione di variabili, 29
espressioni logiche, 24
join di tabelle, 23
lettura, 18, 21
Microsoft Access, 18
proposizione Where, 24
query di parametri, 24, 26
Richiedi valore, 26
salvataggio, 52
salvataggio di query, 31
selezione di campi dei dati, 21
selezione di una sorgente dati, 18
sintassi SQL, 31
sostituzione dei valori nei campi esistenti, 58
sostituzione di tabella, 62
verifica dei risultati, 31
dati categoriali, 111
conversione di dati di intervallo in categorie discrete, 127
dati delimitati da spazi, 32
dati di serie storiche
creazione di nuove variabili di serie storica, 171
definizione di variabili di data, 170
funzioni di trasformazione, 172
sostituzione di valori mancanti, 174
trasformazioni di dati, 169
dati Dimensions, 41
salvataggio, 64

dati pesati, 214
  e file di dati ristrutturati, 214
Definisci insiemi a risposta multipla, 467
  categorie, 467
  dicotomie, 467
  etichette degli insiemi, 467
  nomi degli insiemi, 467
definizione di variabili, 79, 81, 83–87, 108
  applicazione di un dizionario dati, 117
  copiare e incollare attributi, 86–87
  etichette dei valori, 84, 108
  etichette di variabili, 83
  modelli, 86–87
  tipi di dati, 81
  valori mancanti, 85
dendrogrammi
  in cluster gerarchica, 440
Descrittive, 294
  funzioni aggiuntive del comando, 296
  ordine di visualizzazione, 295
  salvataggio di punteggi *z*, 294
  statistiche, 295
descrittive
  in analisi Cluster TwoStep, 434
  in GLM univariato, 349
deviazione media assoluta (AAD)
  in Statistiche di rapporto, 500
deviazione standard
  in cubi OLAP, 321
  in descrittive, 295
  in esplora, 299
  in frequenze, 290
  in GLM univariato, 349
  in medie, 317
  in Report: Riepiloghi per colonne, 482
  in Report: Riepiloghi per righe, 477
  in Riassumi, 312
  in Statistiche di rapporto, 500
di scala, 80
  in analisi di affidabilità, 487
  in scaling multidimensionale, 492
  livello di misurazione, 80
diagnostica di collinearità
  in regressione lineare, 369
diagnostiche per casi
  in regressione lineare, 369
didascalie, 259–260
DiffAdatt
  in regressione lineare, 367
differenza, 172
differenza in beta
  in regressione lineare, 367
differenza meno significativa
  in ANOVA univariata, 335
differenza meno significativa (LSD)
  in GLM, 345
differenza meno significativa (LSD) di Fisher
  in GLM, 345
differenza significativa di Tukey
  in ANOVA univariata, 335
  in GLM, 345
differenza stagionale, 172
differenze tra gruppi
  in cubi OLAP, 324
differenze tra variabili
  in cubi OLAP, 324
differenziale di prezzo (PRD)
  in Statistiche di rapporto, 500
dimensioni, 220
  nella vista riassuntiva, 220
directory temporanea, 535
  impostazione della posizione in modalità locale, 535
  variabile di ambiente SPSSTMPDIR, 535
distanza chi-quadrato
  in distanze, 359
distanza City-Block
  in distanze, 359
distanza city-block
  nell'analisi del vicino più vicino, 393
distanza di Chebychev
  in distanze, 359
Distanza di Cook
  in GLM, 347
  in regressione lineare, 367
Distanza di Mahalanobis
  in analisi discriminante, 415
  in regressione lineare, 367
Distanza di Manhattan
  nell'analisi del vicino più vicino, 393
Distanza euclidea
  in distanze, 359
  nell'analisi del vicino più vicino, 393
distanza euclidea quadratica
  in distanze, 359
distanza Minkowski
  in distanze, 359
Distanze, 357
  calcolo delle distanze tra casi, 357
  calcolo delle distanze tra variabili, 357
  esempio, 357
  funzioni aggiuntive del comando, 360
  in cluster gerarchica, 436
  misure di dissimilarità, 359
  misure di similarità, 360
  statistiche, 357
  trasformazione di misure, 359–360
  trasformazione di valori, 359–360
distanze dei vicini più vicini
  nell'analisi del vicino più vicino, 405
distinzione per gruppi, 192
divisione
  divisione tra colonne del report, 483

divisore di finestre
Editor dei dati, 100
divisore di riquadri
Editor dei dati, 100
dizionario, 44
Informazioni sui dati, 282
dizionario dati
applicazione da un altro file, 117
Dunnett (T3)
in ANOVA univariata, 335
in GLM, 345

Editor dei dati, 77, 79, 86, 92–96, 99–101, 543
allineamento, 86
casi filtrati, 99
definizione di variabili, 79
file di dati multipli aperti, 102, 522
inserimento di dati, 92
inserimento di dati non numerici, 93
inserimento di dati numerici, 93
inserimento di nuove variabili, 96
inserimento di nuovi casi, 95
invio di dati ad altre applicazioni, 543
larghezza di colonna, 86
limitazioni relative ai valori, 94
modifica del tipo di dati, 96
modifica di dati, 94–95
opzioni di visualizzazione, 100
spostamento delle variabili, 96
stampa, 101
viste/riquadri multipli, 100
Visualizzazione dati, 77
Visualizzazione variabili, 78
Editor dei grafici, 509
proprietà, 510
editor della sintassi, 272
autocompletamento, 275
codifica colori, 275
impostazione del testo come commento, 279
indicatori visivi comandi, 273
numeri di riga, 273
opzioni, 539
punti di interruzione, 273, 276, 279
segnalibri, 273, 278
editor della sintassi dei comandi, 272
autocompletamento, 275
codifica colori, 275
impostazione del testo come commento, 279
indicatori visivi comandi, 273
numeri di riga, 273
opzioni, 539
punti di interruzione, 273, 276, 279
segnalibri, 273, 278
elenchi di destinazione, 520
elenchi di variabili, 520
riordinamento degli elenchi di destinazione, 520
elenco dei casi, 310
eliminazione all'indietro
in regressione lineare, 363
eliminazione dell'output, 217
eliminazione di comandi EXECUTE multipli nei file di sintassi, 281
equivalenti
nell'analisi del vicino più vicino, 405
errore standard
in descrittive, 295
in esplora, 299
in frequenze, 290
in GLM, 347, 349
nella Curva ROC, 504
errore standard della curtosi
in cubi OLAP, 321
in medie, 317
in Riassumi, 312
errore standard della media
in cubi OLAP, 321
in medie, 317
in Riassumi, 312
errore standard dell'asimmetria
in cubi OLAP, 321
in medie, 317
in Riassumi, 312
esclusione dell'output da Viewer con SGO, 597
Esplora, 298
funzioni aggiuntive del comando, 302
grafici, 300
opzioni, 302
statistiche, 299
trasformazioni di potenza, 301
valori mancanti, 302
esportazione
modelli, 267
esportazione dell'output, 223, 229, 232
formato Excel, 223, 227, 591
formato HTML, 223
formato PDF, 223, 231, 591
Formato PowerPoint, 223
formato testo, 591
formato Word, 223, 226, 591
HTML, 225
SGO, 585
esportazione di dati, 45, 543
aggiunta di comandi di menu per l'esportazione di dati, 543
esportazione di grafici, 223, 234–235, 576
produzione automatica, 576
eta
in medie, 317
in tavole di contingenza, 305
eta-quadrato
in GLM univariato, 349
in medie, 317
etichette, 243
eliminazione, 243

inserimento di etichette di gruppo, 243
vs. nomi di sottotipi in SGO, 590
etichette dei valori, 84, 93, 100, 108, 529
applicazione a più variabili, 113
copia, 113
in file di dati uniti, 186
inserimento di interruzioni di riga, 84
nella vista riassuntiva, 529
nelle tabelle pivot, 529
nell'Editor dei dati, 100
salvataggio nei file di Excel, 45
utilizzo per l'inserimento dei dati, 93
etichette di gruppo, 243
etichette di variabili, 83, 522, 529
in file di dati uniti, 186
inserimento di interruzioni di riga, 84
nella vista riassuntiva, 529
nelle finestre di dialogo, 6, 522
nelle tabelle pivot, 529
EXECUTE (comando)
copiato dalle finestre di dialogo, 281

F multiplo di Ryan-Einot-Gabriel-Welsch
in ANOVA univariata, 335
in GLM, 345
fattore di inflazione della varianza
in regressione lineare, 369
fattorizzazione alfa, 422
fattorizzazione dell'asse principale, 422
fattorizzazione immagine, 422
file, 220
aggiunta di un file di testo al Viewer, 220
apertura, 14
file attivo, 65, 67
creazione di un file attivo temporaneo, 67
file attivo virtuale, 65
inserimento nella cache, 67
file attivo temporaneo, 67
file attivo virtuale, 65
file BMP, 223, 234
esportazione di grafici, 223, 234
file del modello
caricamento di modelli salvati nei punteggi dei dati, 176
file delimitati da tabulazioni, 14–15, 45–46, 48
apertura, 14
lettura dei nomi di variabili, 15
salvataggio, 45–46
scrittura di nomi di variabili, 48
file delimitati da virgole, 32
file di dati, 14–15, 32, 44–46, 51, 67, 73, 199
aggiunta di commenti, 517
apertura, 14–15
capovolgere, 182
Dimensioni, 41
file di dati multipli aperti, 102, 522
informazioni del dizionario, 44
informazioni sul file, 44
miglioramento delle prestazioni per file di grandi dimensioni, 67
mrInterview, 41
protezione, 65
Quancept, 41
Quanvert, 41
ristrutturazione, 199
salvataggio, 45–46
salvataggio dell'output come file di dati SPSS Statistics, 585
salvataggio di sottoinsiemi di variabili, 51
server remoti, 73
testo, 32
trasposizione, 182
file di dati multipli aperti, 102, 522
soppressione, 106
file di dBASE, 14, 17, 45–46
lettura, 14, 17
salvataggio, 45–46
file di Excel, 14, 16, 46, 543
aggiunta di un comando di menu per inviare dati a Excel, 543
apertura, 14, 16
salvataggio, 45–46
salvataggio delle etichette dei valori invece dei valori, 45
file di fogli elettronici, 14–16, 48
apertura, 16
lettura degli intervalli, 15
lettura dei nomi di variabili, 15
scrittura di nomi di variabili, 48
File di Lotus 1-2-3, 14, 45–46, 543
aggiunta di un comando di menu per inviare dati a Lotus, 543
apertura, 14
salvataggio, 45–46
file di sintassi dei comandi, 281
file di sintassi dei comandi Unicode, 281
file EPS, 223, 235
esportazione di grafici, 223, 235
file giornale, 535
file in formato portabile
nomi di variabili, 46
file JPEG, 223, 234
esportazione di grafici, 223, 234
file PNG, 223, 234
esportazione di grafici, 223, 234
file PostScript (encapsulated), 223, 235
esportazione di grafici, 223, 235
file SAS
apertura, 14
salvataggio, 45
file spp
conversione nei file spj, 584
File Stata, 17
apertura, 17
lettura, 14
salvataggio, 45

file SYSTAT, 14
  apertura, 14
file tab delimitati, 32
file TIFF, 235
  esportazione di grafici, 223, 235
finestra attiva, 5
finestra designata, 5
finestre, 4
  finestra attiva, 5
  finestra designata, 5
Finestre dell'Aiuto, 11
finestre di dialogo, 8, 517–518, 522
  controlli, 8
  definizione di insiemi di variabili, 517
  icone delle variabili, 8
  informazioni sulle variabili, 9
  ordine di visualizzazione delle variabili, 522
  riordinamento degli elenchi di destinazione, 520
  selezione di variabili, 8
  sovrapposte, 6
  utilizzo di insiemi di variabili, 518
  visualizzazione di etichette di variabili, 6, 522
  visualizzazione di nomi di variabili, 6, 522
formati di data
  anni a due cifre, 525
formati di input, 83
formati di valuta, 528
formati di valuta personalizzati, 81, 528
formati di visualizzazione, 83
formato CSV
  lettura dei dati, 32
  salvataggio di dati, 46
formato di file di dati SPSS Statistics
  trasmissione dell'output ad un file dati, 591, 597
formato dollaro, 81, 83
formato Excel
  esportazione dell'output, 223, 227, 591
formato file sav
  trasmissione dell'output a un file di dati SPSS Statistics, 591, 597
formato fisso, 32
formato libero, 32
formato numerico, 81, 83
formato PDF
  esportazione dell'output, 591
Formato PowerPoint
  esportazione dell'output, 223
formato punto, 81, 83
formato stringa, 81
formato virgola, 81, 83
formato Word
  esportazione dell'output, 223, 226, 591
  tabelle grandi, 223
formattazione
  colonne nei report, 476
Frequenze, 289
  formati, 293
  grafici, 292
  ordine di visualizzazione, 293
  soppressione di tabelle, 293
  statistiche, 290
frequenze attese
  in Regressione ordinale, 376
frequenze cumulate
  in Regressione ordinale, 376
frequenze dei cluster
  in analisi Cluster TwoStep, 434
frequenze osservate
  in Regressione ordinale, 376
funzioni, 138
  trattamento dei valori mancanti, 138

gamma
  in tavole di contingenza, 305
gamma di Goodman e Kruskal
  in tavole di contingenza, 305
Generatore di finestre di dialogo personalizzate, 549
  anteprima, 557
  apertura dei file delle specifiche delle finestre di dialogo, 558
  applicazione di filtri agli elenchi di variabili, 563
  browser di file, 569
  casella combinata, 564
  casella di controllo, 563
  controllo gruppo di elementi, 567
  controllo numerico, 566
  controllo testo, 565
  controllo testo statico, 566
  elenco origine, 561
  file di Aiuto, 551
  file pacchetto finestra di dialogo personalizzati (spd), 558
  filtro dei tipi di file, 571
  finestre di dialogo personalizzate per i comandi di estensione, 572
  gruppo di caselle di controllo, 568
  gruppo pulsanti di opzione, 567–568
  installazione delle finestre di dialogo, 558
  lista destinazione, 561
  localizzazione di finestre di dialogo e file di Aiuto, 574
  modello di sintassi, 555
  modifica delle finestre di dialogo installate, 558
  posizione menu, 553
  proprietà della finestra di dialogo, 551
  proprietà sotto-finestra, 572
  pulsante sotto-finestra, 571
  regole di layout, 554
  salvataggio delle specifiche delle finestre di dialogo, 558
  voci di elenco caselle combinate, 564
Generatore di grafici, 505
  modelli, 506
gestione del rumore
  in analisi Cluster TwoStep, 431

GET DATA, 65
  confronto con il comando DATA LIST, 65
  confronto con il comando GET CAPTURE, 65
GGRAPH
  conversione di IGRAPH in GGRAPH, 612
giornale di sessione, 535
giustificazione, 218, 524
  output, 218, 524
GLM
  grafici di profilo, 344
  modello, 341
  salvataggio di matrici, 347
  salvataggio di variabili, 347
  somma dei quadrati, 341
  test Post Hoc, 345
GLM univariato, 339, 350
  contrasti, 343
  diagnostici, 349
  medie marginali stimate, 349
  opzioni, 349
  visualizzazione, 349
grafici, 217, 223, 264, 505, 531
  cenni generali, 505
  copia in altre applicazioni, 223
  creazione, 505
  creazione di grafici da tabelle pivot, 264
  dimensione, 513
  esportazione, 223
  etichette dei casi, 380
  Generatore di grafici, 505
  modelli, 513, 531
  nascondere, 217
  nella Curva ROC, 502
  proporzioni, 531
  riquadri con ritorno a capo automatico, 513
  valori mancanti, 513
grafici a barre
  in frequenze, 292
grafici a dispersione
  in regressione lineare, 365
grafici a scatole
  confronto dei livelli del fattore, 300
  confronto di variabili, 300
  in esplora, 300
grafici a stalattite
  in cluster gerarchica, 440
grafici a torta
  in frequenze, 292
grafici dei pesi fattoriali
  in analisi fattoriale, 423
grafici dei residui
  in GLM univariato, 349
grafici di importanza delle variabili
  in analisi Cluster TwoStep, 433
grafici di normalità detrendizzati
  in esplora, 300
grafici di probabilità normale
  in esplora, 300
  in regressione lineare, 365
grafici di profilo
  in GLM, 344
grafici di variabilità vs. densità
  in esplora, 300
  in GLM univariato, 349
grafici interattivi, 223
  copia in altre applicazioni, 223
grafici parziali
  in regressione lineare, 365
grafici ramo-foglia
  in esplora, 300
grafici: opzioni, 531
grafico dello spazio di funzioni
  nell'analisi del vicino più vicino, 402
grafico di importanza
  in analisi Cluster TwoStep, 433
grafico tabella, 264
griglie, 261
  tabelle pivot, 261

H di Kruskal-Wallis
  in test per due campioni indipendenti, 461
Hochberg (GT2)
  in ANOVA univariata, 335
  in GLM, 345
HTML, 223, 225
  esportazione dell'output, 223, 225

ICC. *Vedere* coefficiente di correlazione intraclasse, 489
icone
  nelle finestre di dialogo, 8
identificatori di comando, 589
IGRAPH
  conversione di IGRAPH in GGRAPH, 612
importanza della variabile
  nell'analisi del vicino più vicino, 404
importazione di dati, 14, 18
impostazione della pagina, 237–238
  dimensioni del grafico, 238
  intestazioni e piè di pagina, 237
indice di concentrazione
  in Statistiche di rapporto, 500
individuazione e sostituzione
  documenti viewer, 221
Informazioni sui dati, 282
  output, 284
  statistiche, 287
informazioni sul file, 44
informazioni sulle variabili, 516
iniziale
  in cubi OLAP, 321
  in medie, 317
  in Riassumi, 312

inserimento di dati, 92–93
  non numerici, 93
  numerico, 93
  utilizzo di etichette dei valori, 93
inserimento di etichette di gruppo, 243
inserimento nella cache, 67
  file attivo, 67
insiemi a risposta multipla
  categorie multiple, 114
  definizione, 114
  dicotomie multiple, 114
  Informazioni sui dati, 282
insiemi di dati
  ridenominazione, 105
insiemi di variabili, 517–518
  definizione, 517
  utilizzo, 518
interruzioni di riga
  etichette di variabili e dei valori, 84
interruzioni di tabelle, 263
intervalli di confidenza
  in ANOVA univariata, 337
  in esplora, 299
  in GLM, 343, 349
  in regressione lineare, 369
  in test T per campioni appaiati, 330
  in test T per campioni indipendenti, 328
  in test T per un campione, 332
  nella Curva ROC, 504
  salvataggio in regressione lineare, 367
intervalli di stima
  salvataggio in regressione lineare, 367
  salvataggio in stima di curve, 382
intervallo
  in cubi OLAP, 321
  in descrittive, 295
  in frequenze, 290
  in medie, 317
  in Riassumi, 312
  in Statistiche di rapporto, 500
intervallo multiplo di Ryan-Einot-Gabriel-Welsch
  in ANOVA univariata, 335
  in GLM, 345
intestazioni, 237
istogrammi
  in esplora, 300
  in frequenze, 292
  in regressione lineare, 365
iterazioni
  in analisi fattoriale, 422–423
  nel cluster con metodo delle K-medie, 444

join esterno, 23
join interno, 23

kappa
  in tavole di contingenza, 305
kappa di Cohen
  in tavole di contingenza, 305
KR20
  in analisi di affidabilità, 489
Kuder-Richardson 20 (KR20)
  in analisi di affidabilità, 489

LAG (funzione), 172
lambda
  in tavole di contingenza, 305
lambda di Goodman e Kruskal
  in tavole di contingenza, 305
Lambda di Wilks
  in analisi discriminante, 415
larghezza di colonna, 86, 250, 261, 534
  controllo della larghezza massima, 250
  controllo della larghezza per il testo a capo, 250
  controllo della larghezza predefinita, 534
  nell'Editor dei dati, 86
  tabelle pivot, 261
lingua
  modifica della lingua dell'interfaccia utente, 522
  modifica della lingua dell'output, 522
linguaggio a comandi, 268
livellamento, 172
livellamento T4253H, 172
livello di misurazione, 80, 111
  definizione, 80
  icone nelle finestre di dialogo, 8
login al server, 69

mappa dei quadranti
  nell'analisi del vicino più vicino, 406
massima verosimiglianza
  in analisi fattoriale, 422
massimo
  confronto di colonne del report, 483
  in cubi OLAP, 321
  in descrittive, 295
  in esplora, 299
  in frequenze, 290
  in medie, 317
  in Riassumi, 312
  in Statistiche di rapporto, 500
matrice di correlazione
  in analisi discriminante, 414
  in analisi fattoriale, 419, 421
  in Regressione ordinale, 376
matrice di covarianza
  in analisi discriminante, 414, 416
  in GLM, 347
  in regressione lineare, 369
  in Regressione ordinale, 376
matrice di modelli
  in analisi fattoriale, 419
matrice di trasformazione
  in analisi fattoriale, 419

media
  di più colonne del report, 483
  in ANOVA univariata, 337
  in cubi OLAP, 321
  in descrittive, 295
  in esplora, 299
  in frequenze, 290
  in medie, 317
  in Report: Riepiloghi per colonne, 482
  in Report: Riepiloghi per righe, 477
  in Riassumi, 312
  in Statistiche di rapporto, 500
  sottogruppo, 315, 320
media armonica
  in cubi OLAP, 321
  in medie, 317
  in Riassumi, 312
media geometrica
  in cubi OLAP, 321
  in medie, 317
  in Riassumi, 312
media mobile a priori, 172
media mobile centrata, 172
media pesata
  in Statistiche di rapporto, 500
media trim
  in esplora, 299
mediana
  in cubi OLAP, 321
  in esplora, 299
  in frequenze, 290
  in medie, 317
  in Riassumi, 312
  in Statistiche di rapporto, 500
mediana dei gruppi
  in cubi OLAP, 321
  in medie, 317
  in Riassumi, 312
mediana mobile, 172
Medie, 315
  opzioni, 317
  statistiche, 317
medie di gruppi, 315, 320
medie di sottogruppi, 315, 320
medie marginali stimate
  in GLM univariato, 349
medie osservate
  in GLM univariato, 349
memoria, 522
menu, 543
  personalizzazione, 543
metafile, 223
  esportazione di grafici, 223
metodi di selezione, 262
  selezione di righe e colonne nelle tabelle pivot, 262
Microsoft Access, 18
minimi quadrati generalizzati
  in analisi fattoriale, 422
minimi quadrati non ponderati
  in analisi fattoriale, 422
minimi quadrati ponderati
  in regressione lineare, 362
minimo
  confronto di colonne del report, 483
  in cubi OLAP, 321
  in descrittive, 295
  in esplora, 299
  in frequenze, 290
  in medie, 317
  in Riassumi, 312
  in Statistiche di rapporto, 500
misura della differenza delle misure
  in distanze, 359
misura delle differenze dei modelli
  in distanze, 359
misura di dissimilarità di Lance e Williams, 359
  in distanze, 359
misura di distanza phi-quadrato
  in distanze, 359
misure di dispersione
  in descrittive, 295
  in esplora, 299
  in frequenze, 290
  in Statistiche di rapporto, 500
misure di distanza
  in cluster gerarchica, 438
  in distanze, 359
  nell'analisi del vicino più vicino, 393
misure di distribuzione
  in descrittive, 295
  in frequenze, 290
misure di similarità
  in cluster gerarchica, 438
  in distanze, 360
misure di tendenza centrale
  in esplora, 299
  in frequenze, 290
  in Statistiche di rapporto, 500
moda
  in frequenze, 290
modalità distribuita, 69–70, 73, 75
  accesso a file di dati, 73
  percorsi relativi, 75
  procedure disponibili, 75
modelli, 86–87, 265, 513, 531
  attivazione, 265
  copia, 267
  definizione di variabili, 86–87
  esportazione, 267
  grafici, 513
  interazione, 265
  nei grafici, 531
  proprietà, 266

stampa, 267
utilizzo di un file di dati esterno come modello, 117
Viewer modelli, 265
Modelli di tabelle, 248–249
applicazione, 248
creazione, 249
modelli fattoriali completi
in GLM, 341
modelli personalizzati
in GLM, 341
modello composto
in stima di curve, 382
modello cubico
in stima di curve, 382
modello di crescita
in stima di curve, 382
modello di Guttman
in analisi di affidabilità, 487, 489
modello di posizione
in Regressione ordinale, 377
modello di potenza
in stima di curve, 382
modello di scala
in Regressione ordinale, 378
modello esponenziale
in stima di curve, 382
modello inverso
in stima di curve, 382
modello lineare
in stima di curve, 382
modello logaritmico
in stima di curve, 382
modello logistico
in stima di curve, 382
modello parallelo
in analisi di affidabilità, 487, 489
modello parallelo esatto
in analisi di affidabilità, 487, 489
modello quadratico
in stima di curve, 382
modello S
in stima di curve, 382
modifica di dati, 94–95
moltiplicazione
moltiplicazione tra colonne del report, 483
mostrare, 217, 247, 260, 544
annotazioni, 260
barre degli strumenti, 544
didascalie, 260
etichette delle dimensioni, 247
righe o colonne, 247
risultati, 217
titoli, 248
mrInterview, 41

nascondere, 217, 247, 260, 544
annotazioni, 260
barre degli strumenti, 544
didascalie, 260
etichette delle dimensioni, 247
righe e colonne, 247
risultati della procedura, 217
titoli, 248
nascondere (escludere) l'output da Viewer con SGO, 597
Newman-Keuls
in GLM, 345
nomi di variabili, 79, 522
file in formato portabile, 46
generati da SGO, 604
nelle finestre di dialogo, 6, 522
nomi di variabili a casi misti, 79
nomi lunghi di variabili con testo a capo nell'output, 79
regole, 79
troncamento dei nomi di variabili lunghe nelle versioni precedenti, 46
nominale, 80
livello di misurazione, 80, 111
notazione scientifica, 81, 522
soppressione nell'output, 522
novità
versione 17.0, 2
numerazione delle pagine, 238
nei report di riepilogo per colonne, 485
nei report di riepilogo per righe, 478
numeri di porta, 70
numero di casi
in cubi OLAP, 321
in medie, 317
in Riassumi, 312

opzioni, 522, 524–525, 528–529, 531, 534–535, 537, 539
anni a due cifre, 525
aspetto della tabella pivot, 534
assegnazione multipla, 541
dati, 525
directory temporanea, 535
editor della sintassi, 539
etichette di output, 529
generali, 522
grafici, 531
script, 537
valuta, 528
Viewer, 524
Visualizzazione variabili, 527
opzioni della riga di comando, 583
Sessioni Utilità SPSS Production, 583
ordinale, 80
livello di misurazione, 80, 111
ordinamento
sovrapposte, 181
ordinamento dei casi, 180
ordinamento delle variabili, 181
ordine di visualizzazione, 242

ortografia, 92
  dizionario, 525
outliers
  in analisi Cluster TwoStep, 431
  in esplora, 299
  in regressione lineare, 365
output, 216–218, 222–223, 239, 524
  allineamento, 218, 524
  centratura, 218, 524
  copia, 217
  copia in altre applicazioni, 223
  eliminazione, 217
  esportazione, 223
  incollare in altre applicazioni, 222
  modifica della lingua dell'output, 522
  mostrare, 217
  nascondere, 217
  salvataggio, 239
  spostamento, 217
  Viewer, 216
output dei grafici, 223
OXML, 610

passaggi fondamentali, 9
PDF
  esportazione dell'output, 223, 231
percentili
  in esplora, 299
  in frequenze, 290
percentuali
  in tavole di contingenza, 308
percentuali complessive
  in tavole di contingenza, 308
percentuali di colonna
  in tavole di contingenza, 308
percentuali di riga
  in tavole di contingenza, 308
percorsi file
  controllo dei percorsi file predefiniti, 535
percorsi file predefiniti, 535
peso di casi, 197
  pesi frazionari nelle tavole di contingenza, 197
phi
  in tavole di contingenza, 305
piè di pagina, 237
pivoting
  controllo con l'output esportato di SGO, 602
PLUM
  in Regressione ordinale, 373
PowerPoint, 229
  esportazione dell'output come PowerPoint, 229
prestazioni, 67
  copia di dati nella cache, 67
previsione
  in stima di curve, 382
produzione automatica, 576
profondità struttura
  in analisi Cluster TwoStep, 431
programmazione con il linguaggio a comandi, 268
proporzioni, 531
proprietà, 250
  sovrapposte, 250
  tabelle pivot, 250
proprietà delle celle, 257–258
pulsante Aiuto, 8
Pulsante di annullamento, 8
Pulsante di ripristino, 8
pulsante Incolla, 8
Pulsante OK, 8
punteggi esponenziali, 152
punteggi fattoriali, 424
punteggi fattoriali di Anderson-Rubin, 424
punteggi fattoriali di Bartlett, 424
punteggi normali
  in classifica casi, 152
punteggi *z*
  in classifica casi, 152
  in descrittive, 294
  salvataggio come variabili, 294
punteggio
  caricamento di modelli salvati, 176
  modelli supportati per l'esportazione e il calcolo del punteggio, 175
  visualizzazione modelli caricati, 179

Q di Cochran
  in test per diversi campioni dipendenti, 465
Quancept, 41
Quanvert, 41
quartili
  in frequenze, 290

R *multiplo*
  in regressione lineare, 369
R-E-G-W *F*
  in ANOVA univariata, 335
  in GLM, 345
R-E-G-W *Q*
  in ANOVA univariata, 335
  in GLM, 345
R^2
  in medie, 317
  in regressione lineare, 369
  modifica di R^2, 369
R^2 corretto
  in regressione lineare, 369
R^2 di Cox and Snell
  in Regressione ordinale, 376
R^2 di Nagelkerke
  in Regressione ordinale, 376
R^2di McFadden
  in Regressione ordinale, 376

raggruppamento
  scelta di una procedura, 427
raggruppamento di righe o colonne, 243
ramificazioni massime
  in analisi Cluster TwoStep, 431
rapporto di covarianza
  in regressione lineare, 367
regressione
  grafici, 365
  Regressione lineare, 362
  regressione multipla, 362
Regressione lineare, 362
  blocchi, 362
  esportazione di informazioni dei modelli, 367
  funzioni aggiuntive del comando, 372
  grafici, 365
  metodi di selezione delle variabili, 363, 371
  pesi, 362
  residuo, 367
  salvataggio di nuove variabili, 367
  statistiche, 369
  valori mancanti, 371
  variabile di selezione, 365
regressione multipla
  in regressione lineare, 362
Regressione ordinale, 373
  collegamento, 374
  funzioni aggiuntive del comando, 379
  modello di posizione, 377
  modello di scala, 378
  opzioni, 374
  statistiche, 373
Regressione parziale minimi quadrati, 384
  esportazione di variabili, 387
  modello, 386
report
  colonne totale, 483
  confronto di colonne, 483
  divisione di valori di colonne, 483
  moltiplicazione di valori di colonne, 483
  report di riepilogo per colonne, 480
  report di riepilogo per righe, 474
  totali composti, 483
Report : Riepiloghi per righe, 474
  colonne di dati, 474
  colonne di separazione, 474
  controllo pagina, 477
  formato colonne, 476
  funzioni aggiuntive del comando, 485
  layout di pagina, 479
  numerazione delle pagine, 478
  ordinamento delle sequenze, 474
  piè di pagina, 480
  spaziatura separazioni, 477
  titoli, 480
  valori mancanti, 478
  variabili nei titoli, 480
report di riepilogo per colonne, 480
Report: Riepiloghi per colonne, 480
  colonne totale, 483
  controllo pagina, 484
  formato colonne, 476
  funzioni aggiuntive del comando, 485
  layout di pagina, 479
  numerazione delle pagine, 485
  totale finale, 485
  totali parziali, 484
  valori mancanti, 485
residui
  in tavole di contingenza, 308
residui cancellati
  in GLM, 347
  in regressione lineare, 367
residui di Pearson
  in Regressione ordinale, 376
residui non standardizzati
  in GLM, 347
residui standardizzati
  in GLM, 347
  in regressione lineare, 367
residui studentizzati
  in regressione lineare, 367
residuo
  salvataggio in regressione lineare, 367
  salvataggio in stima di curve, 382
rho
  in correlazioni bivariate, 351
  in tavole di contingenza, 305
Riassumi, 310
  opzioni, 312
  statistiche, 312
ricerca e sostituzione
  documenti viewer, 221
ricodifica di valori, 127, 143–144, 146–148
ridenominazione degli insiemi di dati, 105
riepilogo degli errori
  nell'analisi del vicino più vicino, 410
righe, 262
  selezione nelle tabelle pivot, 262
rimozione di etichette di gruppo, 243
riordinamento di righe e colonne, 242
rischio
  in tavole di contingenza, 305
Risposte multiple
  funzioni aggiuntive del comando, 473
Risposte multiple: Frequenze, 468
  valori mancanti, 468
Risposte multiple: Tavole di contingenza, 470
  associazione di variabili negli insiemi a risposta multipla, 472
  definizione degli intervalli dei valori, 471
  percentuali basate sui casi, 472
  percentuali basate sulle risposte, 472
  percentuali nelle celle, 472

valori mancanti, 472
ristrutturazione dei dati, 198–199, 201–203, 205–210, 212, 214
cenni generali, 198
creazione di più variabili indice per variabili in casi, 208
creazione di una variabile indice per variabili in casi, 207
creazione di variabili indice per variabili in casi, 205
e dati pesati, 214
esempio di casi in variabili, 202
esempio di due indici per variabili in casi, 207
esempio di indice singolo per variabili in casi, 206
esempio di variabili in casi, 201
gruppi di variabili per variabili in casi, 202
opzioni per casi in variabili, 212
opzioni per variabili in casi, 209
ordinamento di dati per casi in variabili, 212
selezione di dati correnti per variabili in casi, 203
selezione di dati per casi in variabili, 210
tipi di ristrutturazione, 199
ritardo, 142
rotazione di etichette, 243
rotazione equamax
in analisi fattoriale, 423
rotazione oblimin diretta
in analisi fattoriale, 423
rotazione quartimax
in analisi fattoriale, 423
rotazione varimax
in analisi fattoriale, 423

s-stress
in scaling multidimensionale, 492
salvataggio dell'output, 223, 229, 232
formato Excel, 223, 227
formato HTML, 223
formato PDF, 223, 231
Formato PowerPoint, 223, 229
formato testo, 223, 232
formato Word, 223, 226
HTML, 223, 225
salvataggio di file, 45–46
controllo dei percorsi file predefiniti, 535
file di dati, 45–46
file di dati SPSS Statistics, 45
query di file di database, 31
salvataggio di grafici, 223, 234–235
file BMP, 223, 234
file EMF, 223
file EPS, 223, 235
file JPEG, 223, 234
file PICT, 223
file PNG, 234
file PostScript, 235
file TIFF, 235
metafile, 223
scala
livello di misurazione, 111
scaling
tabelle pivot, 250, 255
Scaling multidimensionale, 492
condizionalità, 495
creazione di matrici delle distanze, 494
criteri, 496
definizione della forma dei dati, 494
dimensioni, 495
esempio, 492
funzioni aggiuntive del comando, 496
livelli di misurazione, 495
misure di distanza, 494
modelli di scaling, 495
opzioni di visualizzazione, 496
statistiche, 492
trasformazione di valori, 494
scomposizione gerarchica, 342
script, 543, 547
aggiunta a menu, 543
esecuzione con i pulsanti della barra degli strumenti, 547
linguaggio predefinito, 537
script automatici, 537
segmentazione, 127
Segmentazione grafica, 127
seleziona casi, 193
selezione casi
campione casuale, 196
in base a criteri di selezione, 195
intervallo di casi, 196
intervallo di dati, 196
intervallo di orari, 196
selezione delle funzioni
nell'analisi del vicino più vicino, 407
selezione di casi, 193
selezione in avanti
in regressione lineare, 363
nell'analisi del vicino più vicino, 395
selezione k
nell'analisi del vicino più vicino, 408
selezione k e selezione delle funzioni
nell'analisi del vicino più vicino, 409
selezione per passi
in regressione lineare, 363
seme dei numeri casuali, 139
server, 69–70
aggiunta, 70
login, 69
modifica, 70
nomi, 70
numeri di porta, 70
server remoti, 69–70, 73, 75
accesso a file di dati, 73
aggiunta, 70
login, 69
modifica, 70
percorsi relativi, 75
procedure disponibili, 75

Sessioni Utilità SPSS Production, 576, 580, 582–583
  Conversione dei file della sessione Utilità SPSS Production, 584
  esecuzione di più sessioni Utilità SPSS Production, 583
  esportazione di grafici, 576
  file di output, 576
  opzioni della riga di comando, 583
  pianificazione di sessioni dell'utilità SPSS Production, 583
  regole sintassi, 576
  sostituzione dei valori nei file di sintassi, 580
SGO, 585, 608
  controllo delle tabelle pivot, 591, 602
  esclusione dell'output da Viewer, 597
  formato di file di dati SPSS Statistics, 591, 597
  formato Excel, 591
  formato file sav, 591, 597
  formato PDF, 591
  formato testo, 591
  formato Word, 591
  identificatori di comando, 589
  nomi di variabili nei file SAV, 604
  sottotipi di tabelle, 589
  tipi di oggetti di output, 587
  utilizzo di XSLT con OXML, 610
  XML, 591, 605
sintassi, 268, 279, 546, 576
  accesso a *Command Syntax Reference*, 11
  esecuzione, 279
  esecuzione di sintassi dei comandi con i pulsanti della barra degli strumenti, 546
  file di sintassi dei comandi Unicode, 281
  file giornale, 281
  incollare, 270
  registro di output, 271
  regole delle sessioni Utilità SPSS Production, 576
  regole sintassi, 268
sintassi dei comandi, 268, 279, 543, 547, 576
  accesso a *Command Syntax Reference*, 11
  aggiunta a menu, 543
  esecuzione, 279
  esecuzione con i pulsanti della barra degli strumenti, 547
  file giornale, 281
  incollare, 270
  registro di output, 271
  regole delle sessioni Utilità SPSS Production, 576
  regole sintassi, 268
Sistema di gestione dell'output (SGO), 585, 608
sistema di misurazione, 522
soglia iniziale
  in analisi Cluster TwoStep, 431
somma
  in cubi OLAP, 321
  in descrittive, 295
  in frequenze, 290
  in medie, 317
  in Riassumi, 312
somma cumulata, 172
somma dei quadrati, 342
  in GLM, 341
sostituzione di valori mancanti
  interpolazione lineare, 175
  media della serie., 175
  media di punti vicini., 175
  mediana di punti vicini, 175
  trend lineare, 175
sottoinsiemi di casi
  campione casuale, 196
  selezione, 193, 195–196
sottotipi, 589
  vs. etichette, 590
sottotipi di tabelle, 589
  vs. etichette, 590
sottotitoli
  grafici, 512
sovrapposte, 8, 79, 96, 199, 263, 516–517, 522
  allineamento, 258
  caratteri, 257
  colore di sfondo, 257
  controllo delle interruzioni di tabella, 263
  conversione della sintassi del comando TABLES in CTABLES, 612
  definizione, 79
  definizione di insiemi di variabili, 517
  informazioni sulla definizione, 516
  informazioni sulle variabili nelle finestre di dialogo, 9
  inserimento di nuove variabili, 96
  margini, 258
  nelle finestre di dialogo, 6
  ordinamento, 181
  ordine di visualizzazione nelle finestre di dialogo, 522
  proprietà delle celle, 257–258
  ricerca nell'Editor dei dati, 97
  ricodifica, 143–144, 146–148
  ristrutturazione in casi, 199
  selezione nelle finestre di dialogo, 8
  spostamento, 96
spazio su disco, 65, 67
  temporaneo, 65, 67
spazio su disco temporaneo, 65, 67
Sposta valori, 142
spostamento di righe e colonne, 242
stampa, 101, 235–238, 250, 255, 263
  anteprima di stampa, 236
  controllo delle interruzioni di tabella, 263
  dati, 101
  dimensioni del grafico, 238
  grafici, 235
  intestazioni e piè di pagina, 237
  modelli, 267
  numeri di pagina, 238
  output testuale, 235
  ridimensionamento di tabelle, 250, 255
  spaziatura tra elementi dell'output, 238

strati, 235, 250, 255
tabelle pivot, 235
standardizzazione
in analisi Cluster TwoStep, 431
standardizzazione di valori
in descrittive, 294
statistica di Brown-Forsythe
in ANOVA univariata, 337
statistica di Cochran
in tavole di contingenza, 305
statistica di Mantel-Haenszel
in tavole di contingenza, 305
statistica di Welch
in ANOVA univariata, 337
statistiche descrittive
in descrittive, 294
in esplora, 299
in frequenze, 290
in Riassumi, 312
in Statistiche di rapporto, 500
Statistiche di rapporto, 498
statistiche, 500
statistiche Durbin-Watson
in regressione lineare, 369
statistiche *R*
in medie, 317
in regressione lineare, 369
Statistics Coach, 10
Stima di curve, 380
analisi della varianza, 380
inclusione di costanti, 380
modelli, 382
previsione, 382
salvataggio degli intervalli di stima, 382
salvataggio dei residui, 382
salvataggio di valori attesi, 382
stimatore di Tuckey a doppio peso
in esplora, 299
stimatore M di Andrew
in esplora, 299
Stimatore M di Hampel
in esplora, 299
stimatore M di Huber
in esplora, 299
Stimatori M
in esplora, 299
stime dei parametri
in GLM univariato, 349
in Regressione ordinale, 376
stime di Blom, 152
stime di proporzione
in classifica casi, 152
stime di Rankit, 152
stime di Tukey, 152
stime di Van der Waerden, 152
stime effetto-dimensioni
in GLM univariato, 349
stime potenza
in GLM univariato, 349
strati, 235, 244, 246, 250, 255
creazione, 244
in tavole di contingenza, 304
nelle tabelle pivot, 244
stampa, 235, 250, 255
visualizzazione, 244, 246
stress
in scaling multidimensionale, 492
Student-Newman-Keuls
in ANOVA univariata, 335
in GLM, 345
studio di casi di controllo
T per campioni appaiati, 329
studio di confronti tra coppie
in test T per campioni appaiati, 329
suddivisione di tabelle, 263
controllo delle interruzioni di tabella, 263

T per campioni appaiati, 329
opzioni, 330
selezione di variabili appaiate, 329
valori mancanti, 330
T per campioni indipendenti, 326
definizione dei gruppi, 328
intervalli di confidenza, 328
opzioni, 328
valori mancanti, 328
variabili di raggruppamento, 328
variabili stringa, 328
tabella con chiavi, 186
tabella di classificazione
nell'analisi del vicino più vicino, 409
tabelle di frequenza
in esplora, 299
in frequenze, 289
tabelle grandi
incolla in Microsoft Word, 222
tabelle personalizzate
conversione della sintassi del comando TABLES in CTABLES, 612
tabelle pivot, 217, 222–223, 235, 241–244, 247–248, 250–252, 254–255, 261–264, 534
adattamento alla pagina, 250, 255
adattamento predefinito della larghezza della colonna, 534
allineamento, 258
annotazioni, 259–260
aspetto predefinito per le nuove tabelle, 534
bordi, 254
caratteri, 257
colore di sfondo, 257
controllo delle interruzioni di tabella, 263
copia in altre applicazioni, 223
creazione di grafici da tabelle, 264
didascalie, 259–260

eliminazione di etichette di gruppo, 243
esportazione come HTML, 223
formati di celle, 252
gestione, 241
griglie, 261
incollare come tabelle, 222
incollare in altre applicazioni, 222
inserimento di etichette di gruppo, 243
larghezza delle celle, 261
margini, 258
modifica, 241
modifica dell'aspetto, 248
modifica dell'ordine di visualizzazione, 242
nascondere, 217
pivoting, 241
proprietà, 250
proprietà delle annotazioni, 251
proprietà delle celle, 257–258
proprietà generali, 250
raggruppamento di righe o colonne, 243
rotazione di etichette, 243
selezione di righe e colonne, 262
separazione di righe o colonne, 243
spostamento di righe e colonne, 242
stampa di strati, 235
stampa di tabelle di grandi dimensioni, 263
strati, 244
testo di prosecuzione, 255
trasposizione di righe e colonne, 243
utilizzo di icone, 241
visualizzazione dei bordi nascosti, 261
visualizzazione delle celle, 247
TABLES
conversione della sintassi del comando TABLES in CTABLES, 612
Tamhane (T2)
in ANOVA univariata, 335
in GLM, 345
tau di Goodman e Kruskal
in tavole di contingenza, 305
tau di Kruskal
in tavole di contingenza, 305
tau-*b*
in tavole di contingenza, 305
Tau-*b* di Kendall
in correlazioni bivariate, 351
in tavole di contingenza, 305
tau-*c*
in tavole di contingenza, 305
tau-*c* di Kendall , 305
in tavole di contingenza, 305
Tavole di contingenza, 303
formati, 309
grafici a barre raggruppati, 305
pesi frazionari, 197
soppressione di tabelle, 303
statistiche, 305
strati, 304
variabili di controllo, 304
visualizzazione cella, 308
tavole di contingenza, 303
in tavole di contingenza, 303
risposta multipla, 470
termini di interazione, 341, 378
test a intervallo multiplo di Duncan
in ANOVA univariata, 335
in GLM, 345
Test binomiale, 450
dicotomie, 450
funzioni aggiuntive del comando, 452
opzioni, 451
statistiche, 451
valori mancanti, 451
test del segno
in test per due campioni dipendenti, 459
test della mediana
in test per due campioni indipendenti, 461
test delle linee parallele
in Regressione ordinale, 376
Test delle reazioni estreme di Moses
in test per due campioni indipendenti, 457
Test delle successioni
funzioni aggiuntive del comando, 454
opzioni, 453
punti di divisione, 452–453
statistiche, 453
valori mancanti, 453
Test delle successioni di Wald-Wolfowitz
in test per due campioni indipendenti, 457
Test di additività di Tukey
in analisi di affidabilità, 487, 489
Test *di* Dunnett
in ANOVA univariata, 335
in GLM, 345
Test di Kolmogorov-Smirnov per un campione, 454
distribuzione di test, 454
funzioni aggiuntive del comando, 456
opzioni, 455
statistiche, 455
valori mancanti, 455
test di Levene
in ANOVA univariata, 337
in esplora, 300
in GLM univariato, 349
test di Lilliefors
in esplora, 300
test di linearità
in medie, 317
test di McNemar
in tavole di contingenza, 305
in test per due campioni dipendenti, 459
test di normalità
in esplora, 300

test di omogeneità della varianza
in ANOVA univariata, 337
in GLM univariato, 349
test di omogeneità marginale
in test per due campioni dipendenti, 459
test di Scheffé
in ANOVA univariata, 335
in GLM, 345
test di sfericità di Bartlett
in analisi fattoriale, 421
test di Shapiro-Wilk
in esplora, 300
Test di Sidak
in ANOVA univariata, 335
in GLM, 345
Test *di* Tukey
in ANOVA univariata, 335
in GLM, 345
Test *di* Waller-Duncan
in ANOVA univariata, 335
in GLM, 345
test di Wilcoxon
in test per due campioni dipendenti, 459
test esatto di Fisher
in tavole di contingenza, 305
test M di Box
in analisi discriminante, 414
test non parametrici
chi-quadrato, 447
Test delle successioni, 452
Test di Kolmogorov-Smirnov per un campione, 454
Test per diversi campioni dipendenti, 464
Test per diversi campioni indipendenti, 461
Test per due campioni dipendenti, 459
Test per due campioni indipendenti, 456
Test per confronti a coppie di Gabriel
in ANOVA univariata, 335
in GLM, 345
Test per confronti a coppie di Games e Howell
in ANOVA univariata, 335
in GLM, 345
Test per diversi campioni dipendenti, 464
funzioni aggiuntive del comando, 465
statistiche, 465
tipi di test, 465
Test per diversi campioni indipendenti, 461
definizione dell'intervallo, 463
funzioni aggiuntive del comando, 464
opzioni, 463
statistiche, 463
tipi di test, 462
valori mancanti, 463
variabili di raggruppamento, 463
Test per due campioni dipendenti, 459
funzioni aggiuntive del comando, 461
opzioni, 461
statistiche, 461
tipi di test, 460
valori mancanti, 461
Test per due campioni indipendenti, 456
definizione dei gruppi, 458
funzioni aggiuntive del comando, 459
opzioni, 458
statistiche, 458
tipi di test, 457
valori mancanti, 458
variabili di raggruppamento, 458
test per l'indipendenza
chi-quadrato, 305
test *t*
in GLM univariato, 349
in test T per campioni appaiati, 329
in test T per campioni indipendenti, 326
in test T per un campione, 331
Test *T* di Student, 326
Test *T* per due campioni
in test T per campioni indipendenti, 326
Test T per un campione, 331
funzioni aggiuntive del comando, 332
intervalli di confidenza, 332
opzioni, 332
valori mancanti, 332
testi di Friedman
in test per diversi campioni dipendenti, 465
testo, 32, 220, 223, 232
aggiunta al Viewer, 220
aggiunta di un file di testo al Viewer, 220
esportazione dell'output come testo, 223, 232, 591
file di dati, 32
testo a capo, 250
controllo della larghezza della colonna per il testo a capo, 250
etichette di variabili e dei valori, 84
testo di etichetta verticale, 243
testo di prosecuzione, 255
per tabelle pivot, 255
test*t* dipendenti
in test T per campioni appaiati, 329
tipi di dati, 81, 83, 96, 528
definizione, 81
formati di input, 83
formati di visualizzazione, 83
modifica, 96
valuta personalizzata, 81, 528
tipi di oggetti di output
in SGO, 587
titoli, 220
aggiunta al Viewer, 220
grafici, 512
in cubi OLAP, 325
tolleranza
in regressione lineare, 369
totali finali
nei report di riepilogo per colonne, 485

totali parziali
  nei report di riepilogo per colonne, 484
$T^{quadrato}$ di Hotelling
  in analisi di affidabilità, 487, 489
trasformazioni di dati, 525
  calcolo di variabili, 135
  classifica di casi, 151
  condizioni, 137
  funzioni, 138
  ricodifica di valori, 143–144, 146–148
  ritardo di esecuzione, 525
  serie storiche, 169, 171
  variabili stringa, 137
trasformazioni di file, 198–199
  aggregazione dei dati, 188
  distinzione per gruppi, 192
  ordinamento dei casi, 180
  peso di casi, 197
  ristrutturazione dei dati, 198–199
  trasposizione di casi e variabili, 182
  unione di file di dati, 183, 186
trasposizione di casi e variabili, 182
trasposizione di righe e colonne, 243

U di Mann-Whitney
  in test per due campioni indipendenti, 457
ultimo
  in cubi OLAP, 321
  in medie, 317
  in Riassumi, 312
Unicode, 14, 45
unione di file di dati
  file con casi diversi, 183
  file con variabili diverse, 186
  informazioni del dizionario, 186
  ridenominazione di variabili, 186
utilità SPSS Production, 522
  conversione dei file in sessioni Utilità SPSS Production, 584
  utilizzo della sintassi dei comandi dal file giornale, 522

V
  in tavole di contingenza, 305
V di Cramér
  in tavole di contingenza, 305
V di Rao
  in analisi discriminante, 415
valori attesi
  salvataggio in regressione lineare, 367
  salvataggio in stima di curve, 382
valori attesi ponderati
  in GLM, 347
valori d'influenza
  in GLM, 347
  in regressione lineare, 367
valori estremi
  in esplora, 299
valori mancanti, 85, 513
  definizione, 85
  grafici, 513
  in analisi fattoriale, 425
  in ANOVA univariata, 337
  in correlazioni bivariate, 353
  in Correlazioni parziali, 355
  in esplora, 302
  in regressione lineare, 371
  in Report: Riepiloghi per righe, 478
  in Risposte multiple: Frequenze, 468
  in Risposte multiple: Tavole di contingenza, 472
  in test di Kolmogorov-Smirnov per un campione, 455
  in test per due campioni dipendenti, 461
  in test per due campioni indipendenti, 458
  in test T per campioni appaiati, 330
  in test T per campioni indipendenti, 328
  in test T per un campione, 332
  nei report di riepilogo per colonne, 485
  nei test per diversi campioni indipendenti, 463
  nel test binomiale, 451
  nel test chi-quadrato, 449
  nel Test delle successioni, 453
  nella Curva ROC, 504
  nell'analisi del vicino più vicino, 400
  nelle funzioni, 138
  sostituzione in dati di serie storiche, 174
  variabili stringa, 85
valori mancanti definiti dall'utente, 85
variabile di ambiente SPSSTMPDIR, 535
variabile di selezione
  in regressione lineare, 365
variabili
  ridenominazione per file di dati uniti, 186
variabili data
  definizione di dati di serie storiche, 170
variabili di ambiente, 535
  SPSSTMPDIR, 535
variabili di controllo
  in tavole di contingenza, 304
variabili di raggruppamento, 199
  creazione, 199
variabili di scala
  categorizzazione per la creazione di variabili categoriali, 127
variabili di separazione
  in Aggrega dati, 188
variabili formato data, 81, 83, 525
  aggiungi o sottrai da variabili data/ora, 154
  crea variabile data/ora da un insieme di variabili, 154
  crea variabile data/ora dalla stringa, 154
  estrai parte della variabile data/ora, 154
variabili stringa, 85, 93
  calcolo di nuove variabili stringa, 137
  divisione delle stringhe lunghe nelle versioni precedenti, 46
  inserimento di dati, 93

nelle finestre di dialogo, 6
ricodifica in interi consecutivi, 148
valori mancanti, 85
varianza
in cubi OLAP, 321
in descrittive, 295
in esplora, 299
in frequenze, 290
in medie, 317
in Report: Riepiloghi per colonne, 482
in Report: Riepiloghi per righe, 477
in Riassumi, 312
velocità, 67
copia di dati nella cache, 67
Viewer, 216–220, 238–239, 524, 529
compressione della vista riassuntiva, 219
eliminazione dell'output, 217
esclusione dei tipi di output con SGO, 597
espansione della vista riassuntiva, 219
finestra dei risultati, 216
modifica dei livelli della vista riassuntiva, 219
modifica del carattere della vista riassuntiva, 220
modifica delle dimensioni della vista riassuntiva, 220
nascondere risultati, 217
opzioni di visualizzazione, 524
salvataggio di documenti, 239
spaziatura tra elementi dell'output, 238
spostamento dell'output, 217
vista riassuntiva, 216, 218
visualizzazione di etichette dei valori, 529
visualizzazione di etichette di variabili, 529
visualizzazione di nomi di variabili, 529
visualizzazione di valori di dati, 529
viewer
individuazione e sostituzione di informazioni, 221
ricerca e sostituzione di informazioni, 221
Viewer modelli, 265
vista del modello
nell'analisi del vicino più vicino, 401
vista riassuntiva, 218–219
compressione, 219
espansione, 219
modifica dei livelli, 219
nel Viewer, 218
viste/riquadri multipli
Editor dei dati, 100
Visualizzazione dati, 77
Visualizzazione variabili, 78
personalizzazione, 91, 527

W di Kendall
in test per diversi campioni dipendenti, 465

XML
output OXML da SGO, 610
salvataggio dell'output come XML, 585
struttura della tabella in OXML, 605
trasmissione dell'output a XML, 591
XSLT
utilizzo con OXML, 610

Z di Kolmogorov-Smirnov
in test di Kolmogorov-Smirnov per un campione, 454
in test per due campioni indipendenti, 457